Bastard Algebra

Nick Seaver
Department of Anthropology, UC Irvine
Intel Science and Technology Center for Social Computing

I. "Must Kinship be Dehumanized by Mock-Algebra?"

In a 1930 essay titled, simply, "Kinship," Bronisław Malinowski wrote about blood and ink:

> Much ink has flowed on the problem of blood—'blood' symbolizing in most human languages, and that not only European, the ties of kinship, that is the ties derived from procreation. 'Blood' almost became discolored out of all recognition in the process. Yet blood will rebel against any tampering, and flow its own way and keep its own colour.

He was concerned with the fluid dynamics of relation: how the blood of affinity courses around the sluices and levees of kinship systems and how the ink of theory saturates the pages of the anthropological literature. By 1930, that literature was "flooded" with formalized kinship theory—large vocabularies of kin terms, diagrams, formulas, algebras, and geometries—and Malinowski worried that anthropologists, borne along on the flood of ink, were leaving the blood of kinship behind.

Where W.H.R. Rivers had enthused in *Melanesian Society* about a future in which kinship studies would "resemble a work on mathematics in which the results will be expressed by symbols, in some cases even in the form of equations," Malinowski was skeptical. "I must frankly confess," he wrote, "that there is not a single account of kinship in which I do not find myself puzzled by some of this spuriously scientific and stilted mathematization of kinship facts."

Kinship, he suggested, was not a matter of mathematics, but "of flesh and blood, the result of sexual passion and maternal affection, of long intimate daily life, and of a host of personal intimate interests." The "bastard algebra" of formalized kinship theory could not hope to capture the lived details of relatedness, and while blood flowed its own way and kept its own color, the ink of kinship theory was at risk of drying out. To capture the "intimate data of family life," Malinowski argued, the study of kinship required "full-blooded descriptions" that flowed more like the relationships in question.

1

Here was a paradigmatic statement of ethnographic ideals that has guided many researchers through to the present: the descriptive work of anthropology should hew close to lived experience, and formalisms that abstract these experiences into mathematical rules or representations are suspect. But as Malinowski's argument for full-blooded description eased its way into our disciplinary common sense, his argument against formalism left a loose end. We had a way to talk about flows of blood, but not flows of ink. How were we to understand the family situation of bastard algebra itself?

II. The Ocean

When your research requires sitting through many presentations and reading many more blog posts about big data, you become familiar with the stock images often used to represent it. A tunnel lined with blue 1s and 0s curves so its end is just out of sight. A wave, also made of blue 1s and 0s, crashes over the words "BIG DATA," set in a questionable script typeface. Under cloudy skies, a businessman stands in a rowboat, looking to the horizon on a choppy sea made—of course—of blue 1s and 0s.

In popular and critical discourse about big data, water seems to be everywhere. For stock illustrators, the ocean appears to communicate big data's scale, formlessness, and resource-richness. Perhaps they have tapped into that feeling of limitless, ego-centric potentiality Freud called "oceanic." Blog posts warn about drowning in a tsunami of data. Marketers and managers worry about the flood, while engineers talk about working on the plumbing. In his infamous paean to big data and "the end of theory" in a 2008 issue of *WIRED* magazine, Chris Anderson suggested that the scientific method itself was being rendered obsolete by "the data deluge."

The ocean, as Stefan Helmreich has taught us, has a way of dissolving our ideas about relatedness. Underwater, marine bacteria escape from their lineages and smuggle genes laterally among themselves. The soggy link between genealogy and classification—the tree of life and the order of things—breaks apart in the hands of marine microbial geneticists, while new informatic techniques for sorting and specifying relations among species emerge. In the material-semiotic clutch of water and information technology, relations and our ideas about them are promiscuously reconfigured in a state that Helmreich describes as "hyperactive kinship."

If formalism once signified arid reductionism, big data's oceanic imaginaries appear to have rehydrated it. The "vast gulf" between mathematical analyses and daily life that vexed Malinowski has been filled by a flood of data—credit card transactions, page views, the location of your cell phone, clicks, likes, and shares. The flood sustains new calculations that look less like bastard algebras and more like full-blooded formalisms,

closer to the descriptive immersiveness of ethnography than to the reductive genealogy of the family tree. As the collection of data quickens and sprawls, the floodwaters threaten to overtake the life they nominally represent: personalization algorithms produce a milieu for our activity online, shaped by and for the collection of data.

Under this sea, a form of hyperactive kinship is at work. Ideas about relationships—between cause and effect, representation and reality, affinity and similarity—dissolve and take new form among big data's devotees. Our daily lives are marked by new links with mathematical formalisms, and new relations are found and forged between all sorts of concepts and data points: search terms and epidemiological forecasts, listening patterns and musical taste, "culture" and the relative frequencies of words in books. A cottage industry has sprung up on the banks of big data, dedicated to sharing the intuitively counterintuitive correlations that spring forth from its algorithmic machinery. Correlations, to borrow a phrase from Marilyn Strathern, are always a surprise.

But, in spite of grand claims to the contrary, correlations are not enough. As big data swells, it rocks the boats of critical scholars from across the humanities and social sciences who problematize the too-easy forging of relations and remind us that, although it has taken on the appearance of an objective environment, big data is suffused with interpretation and subjectivity. Not everything makes it into the data set, and things that do are transformed. The businessman in the rowboat is probably looking for "insight," and producing insight is more like making lures than pulling up fish—a matter of intentions and construction, not the retrieving of objective relations from an obliging sea of data. In business, big data is used to attract and convince, to capture and compel rather than to prove. In spite of its vastness, its apparent completeness, its seeming objectivity and synonymy with the real, big data is still representation and not reality—ink (or bits), not blood. As Hrönn Brynjarsdóttir relates in her ethnography of Icelandic fishermen and their data, portage between oceans aqueous and informatic is a lot of work.

III. Dark Sciences

As Malinowski noted in 1930, our theoretical problems, like our kin, are inherited. Humanist and social scientific worries about the algorithmic epistemology of big data descend from a long lineage of critiques of objectivity, formalism, quantification, and automation. Malinowski's argument against bastard algebra is a revealing moment in this family history, where a set of enduring and overlapping epistemic divides is cast in the idiom of kinship: the distinction between *naturwissenschaft* and *geisteswissenschaft*, the quantitative and the qualitative, or, as Clifford Geertz would later contrast, "an experimental science in search of law" and "an interpretative one in search of meaning."

We can spot a family resemblance between Malinowski's "full-blooded description" and Geertz's "thick description": both are calls for renewed attention to the blood and sex of daily life, posed against formalistic alters. Where full-blooded description countered the bastard algebra of kinship, thick description was a humanistic alternative to what Geertz called "an ethnographic algorithm"—the post-war American formalisms of ethnoscience, componential analysis, and cognitive anthropology. These approaches, exemplified by the work of Ward Goodenough and Charles Frake, typically understood culture as a system of taxonomic or behavioral rules to be formally elicited and analyzed (what kinds of plants are there and what are they good for? where should newlyweds live? how should one ask for a drink?). Echoing Malinowski, Geertz argued that these formalisms were "obscured by appeals to dark sciences," mistaking their own methods for the essential stuff of cultural life and pretending "to understand men without knowing them."

These dalliances with the dark sciences were also anthropology's first engagements with computers as research tools. Influenced by the cognitive revolution, which took on the software/hardware distinction as a model for the relationship of mind and brain, post-war American formalists avidly pursued the analytic opportunities offered by electronic computers. They worked on new algorithms for multi-dimensional scaling, which could produce models of a culture's taxonomic "spaces." They designed new techniques for formally eliciting cultural data (sometimes called "white room ethnography"), aimed at producing standardized material for computerized comparison. The Wenner-Gren Foundation hosted a germinal symposium in 1962 on *The Use of Computers in Anthropology,* with topics ranging from the electromagnetic workings of computer memory to statistical analysis to numerical ethnological taxonomy. In his introduction to the conference proceedings, Dell Hymes wrote that the future of computing in anthropology depended on two things: the ability to formalize ethnographic analysis and the building of large, systematic, shared data sets—in short, "big data."

But it was not to be, at least for the anthropologists. "Whatever happened to ethnoscience?," Roger Keesing asked in 1972. Maybe the formalism that once called itself "the new ethnography" collapsed under blistering critiques from figures like Geertz, Hortense Powdermaker, and Gerald Berreman. Maybe its theoretical foundations gave out. Maybe, Keesing jokes, the formalists "bored themselves to death." In any case, the symbolic, interpretive, and reflexive turns soon spun the enthusiasm for formalism out of sociocultural anthropology. Ongoing work with computers was relegated to the disciplinary margins, and a long-proposed "Computing Unit" in the American Anthropological Association never made it out of committee. As

computational techniques grew in the other social sciences, the methodological sea change computers heralded for anthropology in the 1960s never came to pass, and the details of post-war American formalism are scarcely considered relevant to contemporary anthropological practice, except perhaps as epistemological bogeymen.

Malinowski and Geertz's critiques of formal analyses echo in contemporary critiques of big data. Where Malinowski had his kinship algebra and Geertz his ethnoscience, we have "data science," the redundantly named, behaviorism-inflected extraction of "insights" from large sets of interactional data. As a result of a sustained effort to keep formalism out, anthropologists today encounter data science as something utterly foreign to the practice of ethnography—a venture-funded epistemology that encroaches on our disciplinary territory, operationalizing concepts like "culture" in troublesome ways. We may recognize relatives of data science in nearby disciplines like cognitive psychology, statistics, and quantitative sociology, but from anthropology and ethnography in particular, it seems to be qualitatively distinct.

Given the history recounted so far, it should not be surprising that ethnography and big data appear as neatly opposed methodological moieties: as long as we have been defining ethnography's richness, full-bloodedness, or thickness, we have done it in the negative image of formalism. Our disciplinary understanding of ethnography's strengths is entwined with our understanding of formalism's weaknesses. But where Malinowski and Geertz had few kind words for their formalist counterparts, it is now increasingly common to hear that ethnography is not a competitor with big data, but rather a potential collaborator: In a series of posts on the blog *Ethnography Matters,* for example, neologisms like "thick data" and "long data" (big data's phallogocentric pull is apparently irresistible) offer ways to conceptualize ethnographic research as complementary to big data. Especially in the corporate context, ethnography is tasked with putting flesh on big data's bones, with choosing hypotheses to test, or with validating statistical findings, drawing on standard practices from mixed methods research.

These relationships between methods are, by now, such disciplinary common sense that we might be surprised how much work goes into rehearsing and reinforcing them. Either ethnography and formal analysis compete for jurisdiction over concepts like "culture," or they enter into scripted collaborations with each other, drawing on their apparently complementary strengths. Often, these arguments about how big data and ethnography might get along rehearse claims that have defined ethnography and its relationship to other methods throughout ethnography's entire history. The appearance today that ethnography and big data are meeting for the first time, coming from distinct lineages, has been hard won.

Rather than expending our efforts defending thickness, attacking formalism, or regulating the connections between the two, we might investigate these systems of relating themselves. To put an anthropological spin on it, we could study the kinship of method: How are methodological relationships deemed legitimate or illegitimate? What has the rise of "big data" as a discursive and technical phenomenon meant for the ways methods relate? How do various groups of people—data scientists, ethnographers, managers, advertisers, "users"—themselves partially constituted through these relations, imagine them to work? To talk about the kinship of knowledge, we'll want to draw on our knowledge of kinship.

IV. The Kinship of Method

"In the same breath," Marilyn Strathern reminds us in *Kinship, Law, and the Unexpected*, "English-speakers find it possible to talk about practices to do with making kinship and practices to do with making knowledge." Words like "conception" and "relation" have long tied together the mental and the parental, and in anthropology, a field originally defined through its interest in kinship, this tendency is only heightened. Strathern outlines how ideas about parentage, causality, ownership, and creation have been shuttled back and forth between Euro-American discourses about family (for example, defining the "mental" conception of a child in surrogate pregnancy) and knowledge (for example, deciding the "paternity" of a bit of intellectual property). This is no idle metaphor. As Strathern notes, "much of culture is a fabrication of resemblances, a making sense through indicative continuities": in the Euro-American context that spawned anthropology, ethnography, and big data, we grasp the facts of kinship and the facts of knowledge with the same conceptual tools. This coincidence offers some promising routes of inquiry for making sense of methodological relations in the time of big data.

To elaborate, let's return to Malinowski's term "bastard algebra." In what sense are formal analyses of kinship "bastards"? Malinowski seems to be suggesting that these methods are the illegitimate offspring of two families of inquiry — the mathematical and the anthropological. Mathematicians would not claim this algebra as their own, and "the average anthropologist," as Malinowski writes, would not either. To this day, many of us remain uneasy about this mixture, and to borrow another bit of kinship terminology, we tend to establish avoidance relationships with mathematics, when we are not dismissing it outright. Like anyone else, anthropologists are concerned with regulating our kinsmen.

But this raises some questions: How do we decide that an algebra is illegitimate, a bastard? Malinowski's answer—that it is not "full-blooded"—should make us

uncomfortable both for its resonance with the language of scientific racism and for how it lays bare the weakness of ethnographic authority. How do we know that a description is full-blooded (or "thick")? Are such descriptions a distinctive type, or do we find continuous variation between the bastard and the full-blooded, the thick and the thin?

Early studies of kinship emphasized the role that kin systems played in providing an overriding coherence to social life, especially among "primitive" people, but as we've learned from kinship studies since, bastards and other exceptions to the rules are often so prevalent that they draw into question the idea that strict rules govern at all. In *Pul Eliya,* for example, Edmund Leach drew on ethnographic data (and statistical analysis) to argue that the idea of kinship as a set of overriding jural rules falls apart in the face of actual kin practices, which are much more tied to factors like resource availability and the accidents of personal life. The normative, he claimed, is secondary to the normal.

For our kinship of methods, this turn suggests that we set aside the idea that methodological clans are maintained by their intrinsic coherence. Instead of bemoaning the illegitimacy of bastard analysis and trying to purify our bloodlines, we might take our lead from later work in kinship studies and turn our attention to the lived details of methodological relations: the flows of researchers and research money, legitimacy exchanges, and the *imagination* of normative rules of method that only find partial realization in the spread of actual practices.

The "new kinship studies," especially studies of queer kinship and kinship with new reproductive technologies, are suggestive here, drawing together the study of kin practices that are novel and contested with the reflexive study of "kinship" itself as a category. Relatedness, these studies show us, is torqued by technological change, reconceived through concepts like hybridity and networks, and more unstable than the makers of algorithmic co-relations might want to admit. In his introduction to *The Use of Computers in Anthropology,* Dell Hymes wrote that computers introduced "novel relationships, physical, mental, and social" to anthropology. If we want to understand these relationships, we'll have to suspend our investment in idealized methodological clans and instead seek out the local details of how relations among methods are made, unmade, and regulated.

V. The Shore

Imagine yourself suddenly set down on a beach, waves of blue 1s and 0s crashing behind you. Inspired by fanciful images of big data, you had thought the action was under the sea, but here on the shore, spitting sand, you can see the work it takes to get insights out of the ocean. While big data splashes through popular discourse with its oceanic imagery, disruptively washing away old industries, the talk among these folks

is focused on the mundane details of making relations among data points and methods: they want to recommend music based on internet chatter about artists, but the computer has to learn that Justin Bieber and JU$TINNN B3IB3R are the same person; they pipe in metadata from two different companies, but their schemas don't match; somewhere on a server farm, a virtual machine has crashed. This is the real world of data science, and it is simultaneously more boring and more interesting than hype about the world-changing power of data would have us believe.

Here on the shore, we can take what sociologist Celia Lury calls the "frog's-eye view": an amphibious perspective that moves between the watery worlds of big data discourse and the pragmatics of technology in the making. A job posting, for instance, advertises a spot on one of Silicon Valley's many "data teams":

> If you had been born in 800BC you would have been a mystic. You find patterns that others don't see. Maybe you solve disentanglement puzzles in a higher dimensional space without even knowing how you do it.

HR is looking for one of those "new priests and alchemists" Genevieve Bell told us about earlier in this volume, but nobody believes in magic anymore. Or, to put it more anthropologically, nobody realized what belief in magic meant, and as they go about their jobs, they rely on software and studies whose workings they don't fully understand but which they trust to be efficacious. From the history of anthropology and the offhand comments of computer programmers, we know that the boundary between magic and technology is hard to define.

The priests and alchemists may seem like loners, pursuing mathematical truths, but as the Marxist mathematician Dirk Jan Struik reminds us: "Man is a social being even when he worries about the straight lines on hypercones in seven dimensional space," and our mystics only work through their relations with countless others: maintainers of software infrastructure, lawyers, office managers, quality assurance testers, and interns. These characters all have different theories and concerns about what is going on: the marketing intern is enthusiastic about the idea that big data means "the end of theory." The head of the data team tweets, "No data scientist I know believes tripe like that." The old-guard VP thinks the glut of 20-something machine learning enthusiasts are upstart "hipsters," while the office manager calls them "medium T-shirts," after their common torso size. In meetings, they argue and speculate on how to proceed.

With Pierre Lemonnier's *Technological Choices* and Kath Weston's *Families We Choose* under our arm, we might note that the anthropology of technology and the study of kinships both queer and technologically aided have something in common: they share a

concern with choice, flexibility, and legitimate variety in domains where such qualities are often thought to not exist. With our attention cued in this way, we can see how data scientists work as professional bastard-makers, combining data sets, algorithms, and epistemologies in unauthorized ways to produce illicit offspring. This is not the underwater free-for-all of "hyperactive kinship," where it seems like anything might be related to anything else, given enough data, but something more like a tidepool ecosystem, where data, technology, and relationships do not go any which way, but pool and flow along the shore, shaping it in turn. Rather than marvel at the ocean, anthropologists might take to the fractal work of mapping the shoreline.

VI. Bastard Ethnography

The imperative to get off the boat and venture ashore is, of course, classically ethnographic in a Malinowskian sense. Spending time among the bastard algebraists of big data is an excellent way to break down our preconceptions about them. From the shore, we can see the variety of people, epistemologies, and methods that constitute "data science." We can see the countless choices involved in cobbling big data together, moments of ambivalence and constraint, and the thickness of formalism in practice.

As big data's oceanic imaginaries dissolve our common sense about relatedness, it is tempting to imagine that ethnography charts a course back to the "real facts" of relations, as Malinowski called them. If formalist algebras bastardize their sociocultural objects, good old-fashioned ethnography, we suppose, can properly reinstall them in webs of meaning, communities of practice, and social structures. Following the example set by late-twentieth-century laboratory ethnography, this would bring both big data and its objects back under our jurisdiction, as practices that are themselves sociocultural and thus best suited to full-blooded description.

But as I have traced here, the relationship between ethnography and its formalist others is not so clear cut, and our understanding of ethnography's strengths is already predicated on formalism's weaknesses. Instead of taking the encounter with big data as an occasion to rehearse this commonsense relationship between methods, we might take it as an opportunity to examine how methods relate. This investigation would not only give texture and specificity to the practices of big data; it would also draw into question ideas about the coherence and self-evidence of anthropology and ethnography.

Kath Weston has described her encounters with the nostalgic idealism of "real anthropology" — a term wielded by those who long for a simple (and non-existent) past, when the work of anthropology was straightforward and clear, untroubled by native anthropologists, queer and postcolonial research, and the breakdown of distinctions between Self and Other. Like other nostalgias, this one is ironically

forgetful, failing to recall disputes of the past and the shifting boundaries of what has counted as "real anthropology" over time. Such nostalgias forget, for example, the assumption by many formalists that Geertz's interpretive anthropology would be a passing fad, and they leave us unable to recognize the ways that formal methods banished to the past persist into the present. Weston notes the persistence of Euclidean and Newtonian concepts in avant-garde social science metaphor: "borders, lines, intersections, levels, scales, points, grids, and of course the 'trans' that introduces transverse and transept as well as transnational." Alberto Corsín Jiménez has critiqued social anthropology's "algebraic imagination," through which ideas about relations and how to study them are inflected by mathematical formalisms we had thought were long gone.

The persistence of such ideas—the latent algebra in ethnographic theory, the sub rosa geometries of post-structuralism, the defining interdependency of concepts like "thickness" and "algorithms"—reminds us that ethnography is a bastard too, breeding descriptions from illicit encounters, mixing conceptual schemes, and stirring the blood of experience with the ink of theory. As we examine the family situation of bastard algebra, we will have to come to terms with the bastard status of ethnography itself, remaining open to the ubiquity and generative potential of epistemologies that overflow their borders and relate without permission.