Otherness in Focus

Visualizations of Genetic Distances in DNA Studies of Roma

Veronika Lipphardt and Mihai Surdu

Abstract *Our chapter examines how genetic studies postulate and stabilize the existence of homogeneous, easy-to-study groups using labels, images, and sampling decisions. Through two case studies, we show that all these three elements are required to make the truth claim as strong as possible. We begin with a brief review of the literature on visualizations from science and technology studies (STS), art history, image studies, and sociology. Then, in the first case study we give an overview of visualizations in genetic studies of Roma showing how over the years, labels in visualizations have become increasingly abstract and stand in for ever-larger truth claims. We argue that visualizations in DNA studies of Roma contribute to a leap of faith at an ontological level: they stretch argumentative claims from small, unrepresentative samples to entire populations or even continental groups, from familial disease mutations to ethnic disease mutations, and from historical migration routes to narratives of origins and ancestry. In the second case study, we analyze a PCA (principal component analysis) visualization from a genetic study of Roma to problematize the performativity of techno-scientific images in DNA studies of Roma in stabilizing truth claims about the Indian origin of the Roma and their difference from "Europeans." The two-dimensional PCA graph analyzed and the order of data plotting work together to obscure, albeit not quite successfully, the juxtaposition between data of Roma and "Europeans." We demonstrate that the racialization of Roma in the DNA papers we have analyzed occurs in the interplay of sampling, labelling, and visualization. We conclude that visualizations of human diversity are part of simplification work which often conceals biased sampling, the arbitrariness of population labelling and the manipulation of data in the laboratory.*

Introduction

Geneticists who study human population genetics sometimes discuss population labels: if a conventional, widely used label is perceived as insulting to some of the members of the population, a renaming may be considered appropriate. This is the case for "Roma," a name that has been predominantly used in population genetics for roughly thirty years instead of "Gypsies." In 2010, the leading journal of forensic geneticists decided to completely abandon the term "Gypsies" and replace it with "Roma" (Parson and Roewer 2010).

However, both terms fail to describe a population one could study as such. “Roma” does not denote a group bound together by language, culture, tradition, religion, geography, occupation, physical appearance, lifestyle, or self-identification (see, e.g., Surdu and Kovats 2015; Surdu 2016). For many decades, “Gypsies” and “Roma” referred to the label applied to people by others, that is, by non-Roma. Even individuals who identify as Roma would probably not agree to be classified under one population label given the enormous variety of subgroups with distinct identities included under this label. Yet the reason why some geneticists have been eager to study “Roma” as a population for decades was the assumption that they had been genetically isolated from their European surroundings. As they were believed to have come to Europe from India, the idea was that they would display a large proportion of Indian, i.e., non-European, genetic ancestry.

Genetic literature rarely questions the label “Roma,” or whether it denotes a population that can be easily traced and sampled for genetic study. In this paper, we would like to examine three components to demonstrate how genetic studies postulates and stabilizes the existence of a homogeneous, easy-to-study group by means of labels, images, and sampling decisions. While population labels and visual elements are displayed in scientific images to strengthen truth claims about Roma, these images rarely explain the sampling strategies used; rather, these strategies need to be extracted through an in-depth reading of the publication. Yet, as we argue, without specific sampling strategies, these images could never be designed in such a convincing way. All three elements are required to make the truth claim as strong as possible.

We examine two case studies to show how labels stabilize conceptions of Roma groupness in visualizations. In the first case, in which we also give an overview of visualizations in genetic studies of Roma, we demonstrate how over the years, labels in visualizations have become increasingly abstract and stand in for ever-larger truth claims. Second, we analyze a PCA (principal component analysis) visualization from a genetic study in which “Roma” are one of the represented populations and demonstrate how the take-home message extracted from the picture, as it is highlighted in conclusions, title, and abstract, misrepresents the findings.

These two cases are merely two instances of many pictures found in DNA studies on Roma, which result from a vast repertoire of entangled methods, techniques, and visualizations: e.g., linkage maps of chromosomes, single-nucleotide polymorphism (SNP) maps, pedigree structure graphs, phylogenetic trees, electropherograms, allele group frequencies graphs, nonmetric multidimensional scaling (NMDS) graphs, homozygosity maps, scatter plots of short tandem repeat (STR) and SNP profiles, genetic distances maps, and diagrams of median-joining networks. All of these graphics demonstrate the increased diversification in visual strategies initiated at the turn of the twentieth century by physical anthropologists and human geneticists to lend credibility to clustering humans into racial or continental groups (Lipphardt 2015; Lipphardt and Sommer 2015; Sommer 2015).

The publications including the images of our case studies are but a few out of 450 peer-reviewed publications we have collected in a database of genetic studies of Roma, all published between 1921 and 2021. About two-thirds of the articles in this database are DNA-based studies published after 1990 that contain a vast register of visualizations. For analyzing our data material, we depend upon long-standing collaboration and experience with an interdisciplinary group including researchers from the life sciences, mathematics, medicine, and the humanities. This group shares an interest in the genetic data of vulnerable communities, especially of Roma people. In our group, the most fervently discussed topics were visualizations and sampling strategies in DNA studies of Roma. The discussions about epistemic, ethical, and social aspects of DNA studies with Roma have resulted in coauthored publications (Lipphardt, Rappold, and Surdu 2021a; Lipphardt et al. 2021; Lipphardt, Rappold, and Surdu 2021b). The analytical work carried out in our meetings has also been insightful for framing the argument of this article.

Analyzing genetic studies with Roma subjects means taking a perspective from within studies targeting a vulnerable population. While some life scientists argue that the inclusion of minorities and underrepresented groups in global health data collection could allow for a reduction in health inequalities (e.g., Hindorff et al. 2017; Ben-Eghan et al. 2020; Sirugo, Williams, and Tishkoff 2019), scientists also acknowledge the problematic implications of employing ethnic and racializing group labels in medical studies and data collections (e.g., Curtis and Balloux 2020; Ben-Eghan et al. 2020; Lee et al. 2008). Our focus on DNA studies of Roma also has an ELSI (ethical, legal and social implication) dimension which we have addressed in other publications. Roma have a troubled history with human genetics and remain vulnerable today to unethical practices in genetics (Lipphardt, Rappold and Surdu 2021b; Lipphardt et al. 2021). Poor ethical practices also contribute to their misrepresentation in genetic studies (Lipphardt, Rappold, and Surdu 2021a). Finally, our database includes a large enough number of cases to enable an analysis of the style of technoscientific imagery (Bredekamp 2015). For our second case study, we have selected a PCA visualization, as this has become a common form of graphic representation in DNA studies of Roma in the last decade; besides, the article presenting the graph under discussion is a good example of sampling issues we encountered in DNA studies of Roma (Lipphardt, Rappold, and Surdu 2021a).

In what follows, we begin with a cursory review of the literature on visualizations from science and technology studies (STS), art history, image studies, and sociology. Next, we provide a short overview of Roma visualizations in genetic and DNA studies, which includes our first case study, and then continue with an in-depth case study. At the end of the paper, we summarize our argument and draw conclusions.

Visualizations in Science

As Bredekamp et al. (2021[2010], XII) put it, "images … do not merely represent, but veritably *construct*, do not simply illustrate, but actively *bring forth*, that which they show". In scientific work, images are arguments of a *framing* strategy as their producers decide on what aspects are worthy of making visible and what others to leave out of the picture (Lynch and Woolgar, 1990, 2014, Lipphardt and Sommer, 2015). A framing strategy is most obvious in the production of photographs[1], but all types of scientific visualizations use selective depiction to concentrate on what they intend to show.

The role of visualizations in scientific knowledge production can hardly be overestimated, as they are not just illustrations of aspects the text conveys. Many scientists expect visualizations to represent the overall argument of the text so that reading the full text becomes unnecessary. In this case, the captions and inscriptions of the visualizations (in our case, population labels) are crucial for meaning making. In a private conversation with a scientist, he conceded: "The text is just an appendix in our publications. What the people are *reading* in publications in our field are the graphs". This statement validates Goffman's claim that "the picture itself is designed to tell its little story without much textual assistance" (1979, 26).

What is more, images are not just informative but can have aesthetical appeal and can potentially invite the reader to imbue the image with additional layers of meaning. For example, the Icelandic neurologist Kari Stefansson, founder of the biopharmaceutical company deCODE genetics, commented on the deCODE Map of the Human Genome: "This map is to me a thing of beauty. We are looking in quite high definition at the ingenious processes driving the generation of human diversity" (deCODE 2010).

Visualizations in expert publications can also conceal a textual, methodological, or argumentative deficit by making a plea for a specific issue or epistemic object (Surdu 2016). Definitional problems may be resolved through visual stimuli. For example, in twentieth-century physical anthropology, "Gypsies" were pictured as nomads—nomadism being considered the epitome of the group—through images of tents, horses, and traditional clothes. In more recent policy publications, "Roma" —by definition considered as poor—are depicted with symbols of poverty in the background, such as garbage dumps, ruined houses, and unclothed children (Surdu 2016).

In human genetic diversity research, visualizations can also be understood as *obligatory passage points* (Callon 1984) for establishing truth claims on genetic relationships among populations or as a battleground for contestation. Elhaik (2022, 14683) provides

---

[1] What Barthes (1984) has suggested for the functions of photography, namely "to inform, to represent, to surprise, to cause, to signify, to provoke desire" (28), seems to also hold true for scientific visualizations. Far from being an arbitrary choice, the selection of scientific visualizations, especially those used in an ubiquitous way, can indicate the habitus of a scientific field. On the one hand, the image is not merely an expression of the researcher's individual ideas, intentions, or point of view, but, as Bourdieu et al. (1990) argue, also a representation of collective modalities of understanding. On the other hand, what the public is not ideologically prepared to see remains invisible (Sontag 1977, 18).

In science, as elsewhere, photography was considered the paragon of objectivity due to the noninterference of human action and subjectivity: "essentially an act of non-intervention" (Sontag 1977, 11), "an assertion of neutrality" (Sekula 1984, 6), "mechanical objectivity" (Daston and Galison 1992, 82); "the mechanical is here [in photography] a guarantee of objectivity" (Barthes 1980, 278). Yet the objectivity and truth attributed to photography are more of a myth (Goffman 1979; Sekula 1984; Barthes 1980; Bourdieu 1990). In Goffman's words (1979, 20) the myth of photographic objectivity consists in "the very general tendency to confuse realness with representativeness and ideographic with nomothetic validity." The above observations apply not only to photographs but to a wide range of scientific visualizations.

a discussion of recent controversies on human genetics visualizations using PCA scatterplots and the method behind them. The author argues that PCA and its visualizations outcomes in human population genetics are "highly biased" due to "two fallacies: *cherry-picking* or *circular reasoning* (i.e., 'exploration'), the screening and selecting PCA scatterplots that fit preconceived hypotheses while ignoring the other plots" (emphasis in original), thus contesting thousands of findings in genetic studies on human diversity. Lending credibility to visualizations of genetic distances is also part of simplification work resulting from data filtering in scientific practices, as we will demonstrate in our case study (for the general argument see Star, 1983; for the case of data filtering in DNA studies of Roma see Lipphardt, Rappold, and Surdu, 2021a). Filtering out data (through sampling strategies and by excluding "admixture" in the lab) has the effect of making the images' arguments more appealing, convincing, striking, and visible.

This additional layer of simplification resulting from data filtering adds to the work of simplification (Lynch, 1985) and vivification (Lynch, 1988) already embedded in schematic visualizations of which genetic distances are part. Scientific visualizations, particularly relational graphs presenting statistical data, have the role of reducing the complexity of large data sets and of making data easy and quickly readable. Their role is to bring order to an otherwise messy reality (Lynch 1985, 1988). William Playfair, the political economist credited as the founder of statistical graphics, states that his graphs make statistical "studies more clear, and retained more easily by memory" (Playfair 1801, 15, cited in Tufte 2001, 45). Since they are easy to recall by memory, statistical graphics may be also memorable, meaning they may exert a considerable, long-lasting influence on the viewers' perceptions and beliefs. Images are powerful; they exert a fascination on their viewers and can hold them captive (Bredekamp 2021[2010]).

In fact, as Latour (1986) states, inscriptions (including graphs, maps and diagrammatic images) serve a political function in the sociotechnical domain insofar as they aim at "*mobilization*": the aim of winning over as many allies as possible to gain domination over a field. The circulation of inscriptions establishes authorship and originality. Visualizations are part and parcel of persuasion strategies as they are capable of convincing by showing rather than by saying or writing (Latour 1986, 14). Staging, "dramatic visual effects" and oversimplification of what is supposed to be represented are part of the key repertoire of scientific visualizations (Latour 1986, 14).

STS scholars hold that scientific visualizations are intimately related to the truth claims they come with; sometimes they provide *evidence* for invisible phenomena or relationships (Bredekamp 2019; 2021[2010]. These scholars acknowledge that scientific visualizations are not neutral or value-free in the sense of a mechanistic illustration of "data" but are deliberate choices informed by and dependent on scientific methodologies, instruments, theories, concepts, laboratory and field research practices and the assumptions embedded in them (Lynch 1985, 1987; Werner 2015; Bredekamp, Dünkel, and Schneider 2015). Moreover, scientific visualizations are not idiosyncratic, since they derive from and contribute to disciplinary pictorial traditions (Lipphardt, Sommer, and Werner 2015; Bredekamp 2015; Bruhn 2015) and conventions (Bredekamp, Dünkel, and Schneider, 2015).

From an STS perspective, the interest resides in the production and use of scientific visualizations and how they come into being rather than in their correspondence with the objects, phenomena, or relationships depicted by them (Lynch 1988, 2014; Daston 2014; Woolgar 2014).

Various scientific *practices* such as marking specimens, amplifying, creating graphical space and labeling can make objects, relations, and phenomena visible and confer them reality (Hacking 1983; Lynch 1985, 2014). As Hacking (1983) explains (with an example from small particle physics), marking and naming are crucial evidentiary scientific practices: "*if you can spray them then they are real*" (ibid., 23; emphasis in the original). Labeling, which makes visible things that otherwise are invisible or absent (Latour 1986), is also a key operation for the truth claims of scientific visualizations: "Even in scientific illustrations it is the caption which determines the truth of the picture" (Gombrich 1961, 67–68, cited in Goffman 1979, 14). The effects of labeling are ever more important when they apply to human beings, as they contribute to "making up people" (Hacking 1999[1986]).

Our analysis applies the above insights from STS, genetics, and art history to explore the collective apprehension of human genetic diversity resulting from scientific visualizations of genetic distance in DNA studies. By problematizing visualizations of groupness in DNA studies—which are unavoidably racializing in the case of Roma—we attend to the following questions: how are the assumed genetic distances *made*, and how are they made visible? What ontological entities and relationships populate these visualizations? What functions do visualizations of genetic distances have in the epistemic process? How do visualization practices (graphs, diagrams, tables) concur to generate population boundaries? Which aesthetic conventions are used to visualize genetic distances and what role do they play in creating boundaries?

An Overview of Scientific Visualizations in DNA Studies of Roma

Since 2010, population genetic papers on Roma have included technoscientific images in a wide variety of genres. In some cases (e.g., Moorjani et al. 2013; Mendizabal et al. 2012), they occupy half or more of the printed space of the paper. There are photographs, pedigree charts, phylogenetic trees, tables, maps, and PCA scatterplots.

In the illustrations, population labels are used in generalizing ways. For example, pedigree charts appear frequently in medical genetic studies of Roma. One of their functions is to display the analysis units (often extended families) as inbreeding and consanguineous. But many studies using pedigree charts make a great leap by applying labels such as "inbreeding" and "consanguineous" collectively to the Roma population (for some more recent examples, see Gabrikova et al. 2013; García-Magariños et al. 2015; Mašindová et al. 2015; Morar et al. 2011).

Another example, consisting of three publications on the same topic, shows how generalizations are built and transferred in the interplay of sampling, labeling, the production of images and their circulations. In the paper "The Molecular Genetic Basis of Glanzmann's Thrombasthenia in a Gypsy Population in France: Identification of a New Mutation on the Alpha IIb Gene" in the journal *Blood* (Schlegel et al. 1995, 978, figure1), the authors include a pedigree chart titled "Generation tree of gypsy family 1" and refer to a number of forty-two obviously related individuals. Nevertheless, the title of the paper refers to "Thrombasthenia in a Gypsy Population in France" (ibid., 977). The study reports about the discovery of a novel mutation which is called a "Gypsy mutation" (Schlegel et al. 1995, 980, figure 5).

This "Gypsy mutation" pictured by Schlegel et al. (1995) is confirmed and repictured in a genealogical pedigree of a different family by Ruan et al. (1998); the authors name it "gypsy Glanzmann's mutation" in the caption of an image (ibid., 133, figure 3).[2] They call the disease mutation an "abnormality" in the "gene of a European gypsy tribe" (Ruan et al. 1998, 129). Both papers thus take an extended family to stand in for populations that are deemed foreign in their context: "Gypsies" in France, or "Gypsies" in Europe. This exemplifies how a truth claim can be enlarged and extended beyond its actual reach by making a label ever more abstract.

The same disease mutation is discussed in a paper in 2011, where it is called "French Gypsy mutation" (Fiore et al. 2011). The paper's aim is to provide "insights into demographic and population history" (ibid., 983) based on the "French Gypsy mutation." Here, the mutation, which is marked by the label "Gypsy," is no longer the epistemic object of interest; rather, it is an epistemic instrument to find out about another object of interest, namely, the population history. Consequently, the label is black-boxed together with the mutation, and it is taken for granted that it can stand in for the history of a population. "Dating French Gypsy mutation in thrombasthenia" appears as the heading of a pedigree chart presented by Fiore et al. (ibid., 984, figure 2) whose caption contains the label "French Gypsy mutation."

Finally, figure 3 (Fiore et al. 2011, 986) titled "Origin of the French Gypsy Mutation: Origin of Gypsies," illustrates "the migration of Gypsies in France" (984) with colored arrows. In the color-coding for figure 3, pink stands for "European migration to Balkans" (the route from "North India" to "Byzantium empire" [sic]), blue for "migration in Germany" (the route from "Byzantium empire" to "Germany"), green for "introduction of the French Gypsy mutation in France" (the route from "Germany to France") and red (used for three concentric circles and three arrows positioned over "France") for "the founder effect." Superimposed on the physical maps depicting the migration route are color-coded rectangles with time intervals: a pink rectangle with the text "11th century," a blue rectangle with the text "15th century" and a green rectangle with the text "17th–18th." However, the chart only shows the data of one family. It is used to make plausible just one story of how the disease mutation might have reached France and does not provide any context regarding other cases, or possible factors, contributing to the accumulation.

In many papers, visual strategies and labels are used to demonstrate genetic distances between Roma and non-Roma. In these cases, the non-Roma counterpart is given a label that reveals how foreign and distant the Roma are deemed to be. For example, in two tables of a medical paper, the authors use the labels "Gypsies" and "Whites" (Milanov, Topalov, and Kmetski 1999). Table 2 (ibid., 220) reports the prevalence of multiple sclerosis in the city of Sofia ("Gypsies" – n=1 and "Whites" – n=31) and in Samkov ("Gypsies"

---

[2] The name Glanzmann is associated with the disease mutation because the Swiss pediatrician Eduard Glanzmann in 1918 identified (in children from a village in Swiss Alps) and described the rare disease which now carries his name: Glanzmann's Thrombasthenia. See https://www.hog.org/handbook/section/1/glanzmanns-thrombasthenia.

– n=2 and "Whites" – n=15). The title of the paper is "Prevalence of Multiple Sclerosis in Gypsies and Bulgarians," even though the case numbers say little about any population. The biomedical paper Ramal et al. (2004) also compares "Gypsies" and "WCM, white Caucasian Mediterranean" in Tables 3 and 4 (ibid., 937).

Other papers (e.g., Castella et al. 2011, 3762, figure 1) compare "Gypsy patients" and "Caucasian patients," or, in the same paper, synonymously, "patients of the Gypsy ethnic group" and "patients (of white origin)" (Castella et al., 2011, 3760). Again, other papers (such as Martinez-Cruz et al. 2015, 5, figure 2) compare Y-chromosome and mtDNA lineages in "Roma populations" and "Host populations" (with sub-labels "Bulgaria," "Romania," "Hungary," "Ukraine," "Slovakia," and "Spain"). In forensic papers, the comparison group is most often labeled "Caucasian."

These examples reveal that visualizations in DNA studies of Roma perform the function of devices that fix, bring forth and circulate population labels. Visualizations appear as the final outcome of the research process underpinning the theoretical-conceptual research model, recruitment on the field, sampling, labeling, data processing and interpretation. At the same time, they are *vehicles* that carry population labels from the field (usually small, isolated rural communities or institutionalized populations in health units, maternity wards prisons, and armies) to the laboratory and onward to scientific journals and audiences. Visualizations in DNA studies of Roma contribute to a leap of faith at an ontological level: they stretch argumentative claims from small, unrepresentative samples to entire populations or even continental groups, from familial disease mutations to ethnic disease mutations, and from historical migration routes to narratives of origins and ancestry.

Case Study: "Refining the South Asian Origin of the Romani People"

The image we chose to analyze in this paper comes from a genetic study published in 2017, titled "Refining the South Asian Origin of the Romani people" (Melegh et al. 2017). In the abstract, the subsection "Results" reads:

> According to our analyses, Roma showed significant IBD sharing of 0.132 Mb with Northwest Indian ethnic groups. The most significant IBD sharings included ethnic groups of Punjab, Rajasthan and Gujarat states. However, we found also significant IBD sharing of 0.087 Mb with ethnic groups living in Pakistan, such as Balochi, Brahui, Burusho, Kalash, Makrani, Pashtun and Sindhi. (Melegh et al. 2017, 1)

From the title and the abstract, the reader takes away the message that Romani people have predominantly South Asian ancestry, and specifically share a significant portion of their ancestry with populations in Northwest India and Pakistan. There is no mention whatsoever of any West Eurasian or European ancestry of the DNA donors who have been labeled Roma. In the conclusions, the authors go into detail regarding the South Asian ancestry of their probands, while they devote no more than a half sentence to the "West Eurasian ancestry."

The most important method employed in the paper is a visualizing one: PCA. The geneticist Eran Elhaik has recently provided a painstaking critique of this method, as well as a detailed explanation of how it works (Elhaik, 2022). To represent genetic distances, PCAs are used to plot the genetic data of different populations in a two-dimensional space even though the original representation of the data set is three-dimensional. The metric distance within the diagram is meant to denote the genetic distance between the populations. Each colored dot or symbol in the image represents one individual; each color represents one studied population. The populations that are mapped together in Melegh et al. (2017) are the following: Roma, Europeans, Han Chinese, North Indian groups, and South Indian groups. In this case, "Europeans" means, following the sampling logic of the data used in this study, descendants of people who migrated to Utah in past times, where Mormon communities tried to establish a racialized White community.

Notably, readers might easily feel inclined to read historical patterns and movement over time into these images, or to infer a phylogenetic-tree-like relationship between the groups and samples depicted. However, a PCA plot cannot show such causal relationships. It cannot determine how the displayed relationship—whether proximal or distant—between any two sample clusters has come about historically. This issue is up for interpretation, if not for speculation, and typically, there are many different scenarios one could consider.

The two images displayed below are taken from the same graph: image (A) includes the North Indian groups, while image (B) includes the South Indian groups. Any of the group names used for this PCA could be an interesting object of analysis, but here, we focus on "Roma."

In both images, all Roma individuals—represented by red crosses—are closest to the Europeans from Utah, represented by green *x*'s. Were this a three-dimensional representation, one could perhaps find many red crosses hidden behind the green *x*'s; this is because the green *x*'s, as the second plotted group, were plotted on top of the red crosses. In fact, some red dots shine through the green cloud constituted by the green "Europeans," indicating that there is overlap between the two populations. In the logic of PCAs, the population overlap implies that there is no such thing as two distinguishable populations (Elhaik 2022, 14683). Hence, according to this image, Roma, whose symbols are more spread out on the graph than those of the Europeans—who cluster together very closely, as one would expect from a more isolated population—might be understood as a population with a more varied ancestry than Europeans. However, in light of the considerable overlap, Roma and Europeans cannot be distinguished from one another.

Importantly, while the wording of the abstract and the handling of population labels seems to suggest homogeneity among human groups, the graph shows that the group of Roma is not homogeneous, and instead far more scattered than the claim in the abstract suggests. This is also true for the other populations represented in the image, but few of them show a degree of scattering similar to the "Roma" dots. In addition, a population cluster may also look more or less homogeneous depending on the scale at which it is examined: a local population will look very homogeneous when plotted together with populations that have been sampled elsewhere, far away on the globe, while the very same population might look very heterogeneous when plotted together with other local populations from the same region.

One might also ask what is missing in the plot, or what the white spaces between the populations signify. Some North Indian populations in version A are more tightly plotted than any of the South Indian populations in version B. But in both images, there is a big gap between Roma/Europeans and all other populations. This raises the question of what kind of population samples might be able to close the gap, and whether those populations might be more closely related to Roma than any Indian population. If so, one could ask whether the "out-of-India" scenario makes any sense at all, given that there must be alternative scenarios, or whether the most likely scenario can even be inferred from a graph like this, given the variety of other potential scenarios. For example, the red dots labeled "Roma" could show close proximity to samples used to represent Greece or Turkey, which might fall between the dots labeled Roma and Northwest Indian. However, nobody would use this data to infer that Greeks and Turks are more closely related to Northwest Indians than Roma, as it would not resonate with any cultural narrative. A much more obvious interpretation, namely, that Roma seem to be closely related to South East Europeans and have as little to do with Northwest India as Greeks or Turks, would also not stick as it aligns neither with cultural narratives nor with what the authors intend to show.

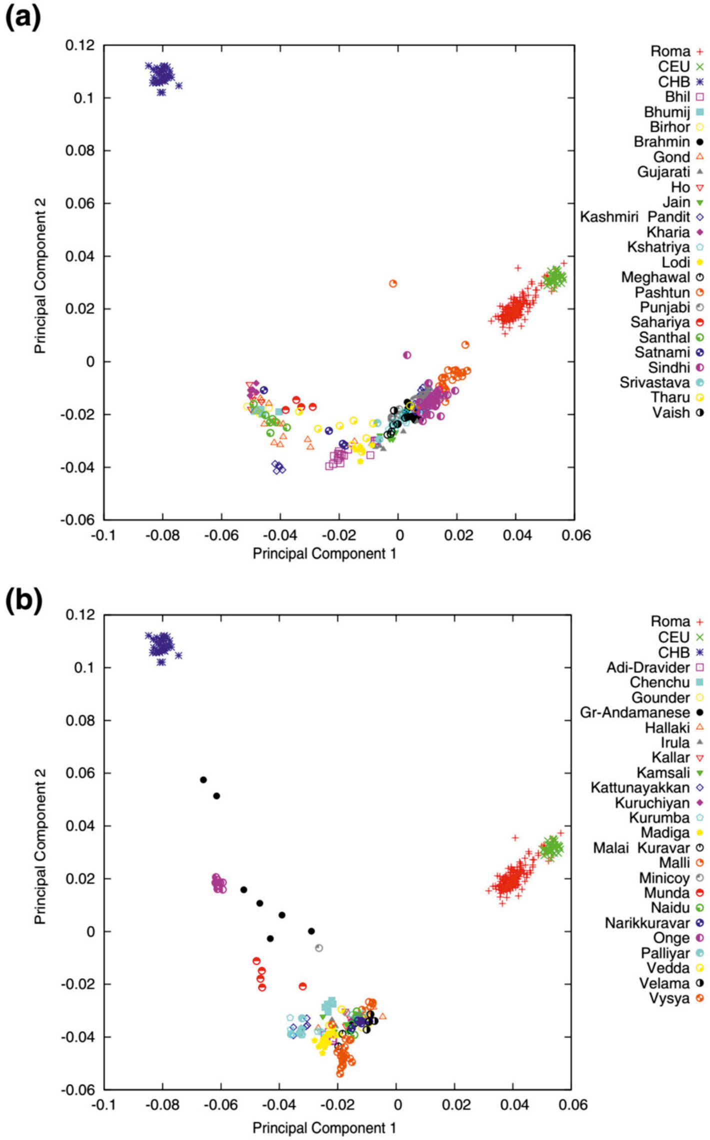
Furthermore, if one reads the image in such a way that any other (unplotted) population between Roma and Northwest Indian were also a candidate for being closely related to Northwest Indians, this would seem counterintuitive, given that the historical movement implied by the image is one of westbound migration of "Roma," and not of a back-and-forth-migrational pattern between other populations, whether close or distant.

The uncertainties multiply, and the data set labeled "Roma" begins to appear even fuzzier when one takes the sampling scheme into consideration. The authors' team states that their sample representing Roma had more individual DNA data than usual: in this case, 158 individuals. These 158 were merged from two different sources. First, 27 samples were taken from Moorjani et al. (2013). Second, another dataset with 152 individual data labeled "Roma" was taken from Mendizabal et al. (2012). Both datasets had been sampled in international collaborations. For the smaller data set, their authors state:

> We collected 27 Roma samples belonging to six groups that were sampled from four countries in Europe from Hungary (3 linguistically and culturally separated sub-groups: 7 samples from Olah (Vlah), 4 samples from Beas (Boyash) and 4 samples from Romungro) [sic], 4 samples from Romania, 4 samples from Spain and 4 samples from Slovakia (Slovakian speaking Roma). (Moorjani et al. 2013, 8)

And as for their sampling criteria, the authors state: "Roma individuals self-reported as being descendants of the same tribe for at least three generations" (ibid).

Figure 1: reproduced from: Melegh et al. 2017. "Refining the South Asian origin of the Romani people." BMC Genetics 18(1), 1–13, fig. 2, 6. The original figure caption reads: "Relationship of Roma to European and South Asian Populations. The principal component analysis results were plotted on the two principal components with the highest eigenvalue. One symbol represents one individual. (a) Relationship of Roma to European and North Indian groups. (b) Relationship of Roma to European and South Indian groups. Separating North and South Indian populations have only practical purpose in order to give a better overview. Both graphs are the result of the same PCA."



CEU - Utah residents with Northern and Western European ancestry from the CEPH collection.
CHB - Han Chinese in Beijing, China.

The second set of data, used in Mendizabal et al. 2012, is much more complex to grasp. 152 individuals from 13 countries were asked to contribute their DNA data. The largest sample comes from Bulgaria (18 individuals), the smallest from Estonia and Wales (7 individuals each). The authors state that "all individuals included in this study were self-identified as Romani" (Mendizabal et al. 2012, Supplemental Information, 1). They also state that "importantly, the self-identification as Romani is a delicate matter in some European countries due to the social stigma attached to Romani identity; hence additional information obtained in sampling can be scant" (ibid). Nonetheless, some information is given for each of the thirteen data sets; most often, the authors state that the people labeled Romani were native speakers of the dominant national language. Some of the datasets had been published before; for others no such information is provided although the authors had made use of them in other studies where publication information is given.

However, not all the samples gathered in this dataset involved self-identification as Romani. Two sub-datasets come from forensic genetic studies (Irwin et al. 2007, Deligiannidis et al. 2006), and one of them—Irwin—lists coauthors from law enforcement institutions. Parts of the data had been collected more than a decade prior to publication. Hence, informed consent was most probably not provided in those studies—and none of the studies mention any self-identification procedure. Self-identification is questionable for at least part of the data, and identification by others—state authorities, for example—can be assumed (for further sampling issues of said data set, see Lipphardt et al. 2021a).

Altogether, the overall sample of 206 individuals seems patchy and incoherent with regard to the sampling scheme. 54 individuals had to be excluded from the sample "for various reasons." The remaining 152 were analyzed and later transferred to the Melegh et al. (2017) study.

Here, another exclusion took place which affected the data represented in the image. Melegh et al. state that they merged the two datasets, that is, 27 and 152 individuals, for a total of 179, but then reduced it to 158. Hence, they subtracted 21 individual datasets, which is more than 10 percent of the overall data. The rationale for excluding these datasets was the following:

Based on PCA and clustering methods, we removed Roma individuals from the merged Roma dataset, which showed significant admixture with non-Roma Europeans. (Melegh et al. 2017, 2)

If an individual dataset showed "significant admixture with non-Roma Europeans" in the lab analysis, it was to be removed from the overall "Roma" dataset. In the twenty-seven individuals from the first source publication, the authors had paid attention to sampling individuals descended from "the same tribe" for at least three generations. It is likely that in the other study (152 individuals) similar sampling strategies were in place. This is already an attempt at filtering out individuals with mixed ancestry in the field before a DNA sample is donated. The second filtering step happens in the lab, in a first PCA run which is not published and is meant to exclude those who have been included in the DNA donor panel when field samples were collected, despite making tribal descendance the criterion for inclusion. Since this first PCA is not published, one cannot ascertain whether other filtering mechanisms have played a role as well. The authors, however, have not only filtered out the Roma sample in the lab but also the sample of Indian populations. As the paper explains in its methodological section, seven Indian groups were removed from the analysis: Siddis because they "have significant recent African related ancestry"; and six other Indian populations because they "showed complex, mainly East Asian related ancestry in the preliminary analyses" (ibid., 2).

Accordingly, the dataset representing Roma has gone through a two-step filtering process to represent individuals who are as distinct as possible from Europeans. Similarly, the Indian data were filtered to remove individuals from the analysis that presented complex, mixed ancestries. Despite these filtering mechanisms, the graph shows no gap between the Roma and Europeans but considerable overlap.

In the results, the authors state: "The Roma are on this cline between the Europeans and South Asians, but are closer to the European samples. Pakistani groups are the closest South Asian groups to the Roma." (ibid., 5) This is not incorrect, but we have noted a clear gap between any South Asian group and Roma, while we have also noted considerable overlap between Roma and Europeans, even though Roma data had been filtered to make them more distinct from Europeans. Thus, the authors make the result of close relatedness or even identical ancestry between Roma and Europeans sound less clear and strong than it is. In addition, in the results, the authors state, "applying F4 Ratio Estimation on this setup, our results showed that Roma have on average 81.08 +/– 0.53 percent West Eurasian related ancestry." (ibid., 5). In the discussion, they state:

> Population structure and ancestry estimation analyses using PCA and model-based clustering methods placed Roma between Europeans and South Asians. These analyses anticipated that European (or more precisely West Eurasian) ancestry in Roma is significant and its proportion is higher than the proportion of the Indian ancestry. (Ibid., 10)

Throughout the discussion, the authors report that all their methods have yielded a strong European ancestry in Roma. But in the conclusions, they do not mention this quantitative aspect when they summarize their findings:

> In conclusion, the results of our study suggest that the West Eurasian ancestry of Roma originates likely from Central and East Europe, and Northwest India plays an even more important role in the South Asian ancestry of Roma than previous studies suggested. Our results also suggest that besides Northwest Indian populations, Pakistani populations play also an important role of the source of South Asian ancestry of Romani people. These new findings extend the South Asian origin of the Romani people making the Pakistani region a similarly important source of ancestry for the Romani people as the Indian subcontinent. (Ibid., 11f.)

This summary does not demonstrate that, according to their analysis, an overwhelming amount of Roma ancestry is European (or, as the authors specify, "more precisely West Eurasian"). The largest part of the summary is dedicated to South Asian ancestry. Considering how the results are presented in the abstract at the beginning of the paper makes this neglecting even more striking: the overwhelming European ancestry of Roma is not even mentioned there – it's only the Asian ancestry that made it to this summarized result section in the abstract on the front page.

With our case study, we aim to demonstrate how labels, sampling, and visualization are interconnected procedures that stabilize truth claims about the Indian origin of the Roma and their difference from "Europeans." Firstly, the research design operates with assumptions about groupness, and assumes the existence of neatly separable groups that fit the labels chosen by the researchers. Our problematization of these assumptions begins on the methodological level with sampling schemes: datasets assembled from various sources through incoherent sampling schemes cannot adequately represent any group. Had the geneticists wanted to show genetic relatedness in a specific social group, they would have had to ensure that all members of the group under study were included under the same sampling criteria. Instead, they lumped together data, including data from people who did *not* self-identify as Roma, which was collected under heterogeneous sampling schemes, and then curated it to make it fit within the confines of a more homogeneous group. We agree that genetic differences exist and can be visualized; however, we contest that these differences easily and clearly align with meaningful social groupings.

Second, no matter whether the grouping is done in/correctly or un/problematically, the forced tracing back of "Roma" ancestry to India, and the active construction of such an ancestry, is at best biased; at worst, it is racist, even if not deliberately so. The sampling strategy proceeds with the deliberate selection only of those individuals who maximize the hypothesized isolation and Indian origin of the Roma (the criterion, "descendants of the same tribe for at least three generations," serves as a proxy to that hypothetical ancestral group); the small sample sizes used by the researchers exclude "admixed" individuals by sampling design. The third step represents the culmination of circular reasoning: in the laboratory, data are further filtered, and those samples of Roma and Indians which show "admixture" are removed from the analysis. In the fourth step, the two-dimensional PCA graph and the order of data plotting work together to obscure, albeit not quite successfully, the juxtaposition between data of Roma and "Europeans." The PCA plot serves to demonstrate that "Roma" are outsiders to Europe and that they are distinct from those who inhabit "Europe" as "natives." The underlying rationale and narrative of the genetic studies on Roma thus supports exclusionist and racist political agendas that are expanding in EU member states today. While the PCA plot is not sufficiently convincing in showing genetic distances between Roma and "Europeans," the interpretation of the graph, which contradicts what is shown, textually accomplishes the final operation in stabilizing the truth claim that Roma are from India.

Conclusions

Our chapter analyzes the forms of thinking and the scientific practices that generate (visualizations of) genetic distances in DNA studies of Roma. The switch from a "representational idiom" to a "performative idiom" (Pickering 1995) and to the scientific practices, a methodological route that we follow, is a well-established research tradition in studies of science (e.g., Hacking 1983; Barad 2003; Latour 2005; Cetina 1999; Law 2004) including the analysis of scientific visualizations (e.g., Burri 2012; Daston 2014; Dumit and de Laet 2014; Latour 1987; Law 2014; Lynch 1985, 1988, 2014) to which we aim to contribute. In our approach, we follow Bredekamp, Dünkel and Schneider (2015, 2) who argue that technoscientific images are "productive agents and distinctive multi-layered elements of the epistemic process."

Visualizations of different sorts are commonplace in DNA studies of Roma; in more recent publications since 2010, they occupy a central role in the economy of scientific publications. The visualizations we have analyzed are part of a trend towards abstraction in pictorial traditions of human variation, which already began in the second half of the twentieth century due to increased sophistication in statistical and graphical techniques. Within this development towards abstraction, the preferred form is that of the diagrammatic tree (Lipphardt and Sommer 2015; Lipphardt 2015). The latest genre within the visual repertoire for picturing human variation is the PCA scatterplot; for our case study, we have chosen a PCA scatterplot to problematize the performativity of technoscientific images in DNA studies of Roma. Within human population genetics, PCA methods and scatterplots are already contested techniques; Elhaik (2022) calls for a reevaluation of between 32,000 and 216,000 genetic studies based on them due to their biased findings. Regardless of whether the field of human population genetics will heed Elhaik's call, the effects produced by these studies—the naturalization of social differences between human groups cemented by labels and cultural narratives—will surely be long-lasting.

As our case study demonstrates, visualizations of human diversity are artifacts that may misrepresent the populations they purport to depict as they often conceal biased sampling, the arbitrariness of population labelling and the manipulation of data in the laboratory. As we have shown, visualizations are part of simplification work carried out by data filtering in the field and in the laboratory resulting in an exaggeration of differences (or distances) among the represented populations. The role of visualizations is to fix an absent reference – populations substituted by labels—and to arrest a dynamic process of continuous mixing among groups, which are instead presented as delimited entities. Visualizations act as vehicles of typological thinking as they operate at an ontological level with continental clusters of populations, and as they present populations as homogenous, discrete entities (hence "types"). Visual strategies and labels are being used to demonstrate genetic distances: at its core, the diversity is both represented and *generated* by genetic distances.

Our contribution aims to critically address the "beyond race" claim in the life sciences and, to use the words of M'charek (2022, 1), "to stay curious about what race is made to be in practice, how it manifests and what politics it does." Our argument is that the *punctum* (in Barthes's 1984 understanding) of past and present visions shared by scientific and expert practices we have analyzed is *race*, understood as a "relational object" created through scientific practices and situated "beyond fact and fiction" (M'charek 2013). The racialization of Roma in DNA papers we have analyzed occurs in the interplay of sampling, labeling, and visualization. Roma are often sampled under the assumption of a consanguineous, inbreeding "genetic isolate"; further on in the laboratory, only those samples deemed to best represent Indian origins are retained, bolstering assumptions related to Roma populations' "purity" and "admixture." Samples gathered and manipulated on essentialist assumptions are then labeled and grouped mostly under continental labels, except for a few cases when racial labels are used (e.g., “Whites” and “Gypsies”). Finally, the interpretation of the results and the reduction of information in the abstract is biased as it overemphasizes the importance of South Asian ancestry. In this way, visualizations can naturalize racial descriptors and racial differences while concealing the ways in which they have been produced.

Finally, we have been asked to constructively propose a more nuanced way of capturing groupness, or of visualizing genetic relationships with PCA images. As a preliminary proposal, we would respond as follows: one could use a large, heterogeneous dataset in a PCA plot. The dataset could include data from individuals who have self-identified clearly and freely using any population label of their choice. Datasets would not be colored or symbolized using population labels; instead, one would print all dots in gray. Next, one could look for clusters of gray dots. If there were clearly identifiable clusters, one could ask whether the annotating data correlated to this cluster had something in common: perhaps self-identification with the same population label, birthplace, or language? If there were clear correlations, one could speak of a social groupness that is also reflected in genetic data. If not, one would have to concede that the relationship between socially ascribed, self-assigned, and genetic groupness is much more complex than the simplistic groupness assumptions within these genetic studies of Roma.

References

Barad, Karen. 2003. “Posthumanist Performativity: Toward an Understanding of How Matter Comes to Matter.” *Signs: Journal of Women in Culture and Society* 28 (3): 801–831.

Barthes, Roland. 1984. *Camera Lucida: Reflections on Photography*. London: Flamingo.

Barthes, Roland. 1980. “Rhetoric of the Image.” In *Classics Essays on Photography*, edited by Alan Trachtenberg, 269–285. New Haven, CT: Leete’s Island Books.

Ben-Eghan, Chief, Rosie Sun, Jose Sergio Hleape et al. 2020. “Don’t Ignore Genetic Data from Minority Populations.” *Nature* 585 (7824): 184–186.

Birch, Kean, and Fabien Muniesa, eds. 2020. *Assetization: Turning Things into Assets in Technoscientific Capitalism*. Cambridge, MA: MIT Press.

Birch, Kean. 2019. *Neoliberal Bio-Economies? Co-Constructing Markets and Natures*. Cham: Palgrave Macmillan.

Bourdieu, Pierre, and Luc Boltanski, eds. 1990. *Photography: A Middle-Brow Art.* Cambridge: Polity Press.

Bredekamp, Horst. 2021. *Image Acts: A Systematic Approach to Visual Agency.* Berlin: de Gruyter.

Bredekamp, Horst. 2019. *Darwin’s Corals: A New Model of Evolution and the Tradition of Natural History*. Berlin: de Gruyter.

Bredekamp, Horst, Vera Dünkel, and Birgit Schneider, eds. 2015. *The Technical Image: A History of Styles in Scientific Imagery*. Chicago: University of Chicago Press.

Bredekamp, Horst. 2015. “A History of Styles of Technical Imagery: Between Description and Interpretation. A Conversation with Horst Bredekamp.” In *The Technical Image*, edited by Horst Bredekamp, Vera Dunkel, and Birgit Schneider. 18–32.

Bredekamp, Horst, Vera Dünkel, and Birgit Schneider, 2015. "Introduction: The Image – A Cultural Technology: A Research Program for a Critical Analysis of Images." In *The Technical Image*, edited by Horst Bredekamp, Vera Dunkel, and Birgit Schneider. 1–8.

Bruhn, Matthias. 2015. "Beyond the Icons of Knowledge: Artistic Styles and the Art History of Scientific Imagery." In *The Technical Image*, edited by Horst Bredekamp, Vera Dunkel, and Birgit Schneider. 36–48.

Burri, Regula Valérie. 2012. "Visual Rationalities: Towards a Sociology of Images." *Current Sociology* 60 (1): 45–60.

Callon, Michel. 1984. "Some Elements of a Sociology of Translation: Domestication of the Scallops and the Fishermen of St. Brieuc Bay." *The Sociological Review* 32 (1 Suppl.): 196–233.

Castella, Maria, Roser Pujol, Elsa Callén et al. 2011. "Origin, Functional Role, and Clinical Impact of Fanconi Anemia FANCA Mutations." *Blood* 117 (14): 3759–69.

Cetina, Karen Knorr. 1999. *Epistemic Cultures: How the Sciences Make Knowledge*. Cambridge, MA: Harvard University Press.

Catelijne Coopmans, Janet Vertesi, Michael E. Lynch, and Steve Woolgar, eds., 2014. In *Representation in Scientific Practice Revisited*, edited by Catelijne Coopmans, Janet Vertesi, Michael E. Lynch, and Steve Woolgar. Cambridge, MA: MIT Press.

Curtis, David, and François Balloux. 2020. "Topical Ethical Issues in the Publication of Human Genetics Research." *Annals of Human Genetics* 84 (4): 313–314.

DeCODE Genetics. 2010. "The New deCODE Map of the Human Genome." Accessed November 20, 2024. www.decode.com/map-of-the-human-genome/

Daston, Lorraine. 2014. "Beyond Representation." In *Representation in Scientific Practice Revisited*, edited by Catelijne Coopmans, Janet Vertesi, Michael E. Lynch and Steve Woolgar. 319–323.

Daston, Lorraine, and Peter Galison. 1992. "The Image of Objectivity." *Representations* 40: 81–128.

Dumit, Joseph, and Marie de Laet. 2014. "Curves to Bodies: The Material Life of Graphs." In *Routledge Handbook of Science, Technology, and Society*, edited by Daniel Lee Kleinman and Kelly Moore, 93–112. London/New York: Routledge.

Elhaik, Eran. 2022. "Principal Component Analyses (PCA)-based Findings in Population Genetic Studies Are Highly Biased and Must Be Reevaluated." *Scientific Reports*, 12 (1): 14683.

Fiore, Mathieu, Xavier Pillois, Paquita Nurden, Alan T Nurden, Frédéric Austerlitz. 2011. "Founder Effect and Estimation of the Age of the French Gypsy Mutation Associated with Glanzmann Thrombasthenia in Manouche Families." *European Journal of Human Genetics* 19 (9): 981–987.

Fleck, Ludwik. 1979. *Genesis and Development of a Scientific Fact*, translated by Frederick Bradley and Thaddeus J. Trenn. Chicago: University of Chicago Press.

Gabrikova, Dana, Martin Mistrik, Jarmila Bernasovska et al. 2013. "Founder Mutations in NDRG1 and HK1 Genes Are Common Causes of Inherited Neuropathies among Roma/Gypsies in Slovakia." *Journal of Applied Genetics* 54 (4): 455–460.

García-Magariños Manuel, Thore Egeland, Ignacio López-de-Ullibarri, Nils L Hjort, Antonio Salas. 2015. "A Parametric Approach to Kinship Hypothesis Testing Using Identity-By-Descent Parameters." Statistical Applications in *Genetics and Molecular Biology* 14 (5): 465–479.

Goffman, Erving. 1979. *Gender Advertisements*. Cambridge, MA: Harvard University Press.

Hacking, Ian. 1983. *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*. Cambridge: Cambridge University Press.

Hacking, Ian. 1999. "Making Up People." In *Reconstructing Individualism: Autonomy, Individuality, and the Self in Western Thought*, edited by Thomas C. Heller, Morton Sosna, and David Wellbery, 222–236. Stanford: Stanford University Press.

Hindorff, Lucia. A., Vence L Bonham, Lawrence C Brodye et al. 2018. "Prioritizing Diversity in Human Genomics Research." *Nature Reviews Genetics* 19 (3): 175–185.

Latour, Bruno. 2005. *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford: Oxford University Press.

Latour, Bruno. 1987. *Science in Action*. Cambridge, MA: Harvard University Press.

Latour, Bruno. 1986. "Visualization and Cognition." *Knowledge and Society: Studies in the Sociology of Culture Past and Present* 6: 1–40.

Law, John. 2014. "Indistinct Perception." In *Representation in Scientific Practice Revisited*, edited by Catelijne Coopmans, Janet Vertesi, Michael E. Lynch and Steve Woolgar. 337–343.

Law, John. 2004. *After Method: Mess in Social Science Research*. London: Routledge.

Lee, Sandra Soo-Jin, Joanna Mountain, Barbara Koenig et al. 2008. "The Ethics of Characterizing Difference: Guiding Principles on Using Racial Categories in Human Genetics." *Genome Biology* 9 (7): 1–4.

Lipphardt, Veronika, Gudrun A. Rappold, and Mihai Surdu. 2023, unpublished. "Ethical standards in forensic genetic research – a critical appraisal of Roma studies."

Lipphardt, Veronika, Gudrun A. Rappold, and Mihai Surdu. 2021. "Representing Vulnerable Populations in Genetic Studies: The Case of Roma." *Science in Context* 34 (1): 69–100.

Lipphardt, Veronika, Mihai Surdu, Nils Ellebrecht, Peter Pfaffelhuber, Matthias Wienroth, and Gudrun A. Rappold. 2021. "Europe's Roma People Are Vulnerable to Poor Practice in Genetics." *Nature* 599: 368–371.

Lipphardt, Veronika, Gudrun A. Rappold, and Mihai Surdu. 2021. "Protect Minorities in Genetic Research." *Science* 373 (6562): 1452–1452.

Lipphardt, Veronika, and Mihai Surdu. 2018, unpublished. Genetic studies on Roma 1921–2018. Internal Report.

Lipphardt, Veronika. 2015. "Traditions and innovations: Visualizations of human variation, c. 1900–38." *History of the Human Sciences* 28 (5): 49–79.

Lipphardt, Veronika, and Marianne Sommer. 2015. "Visibility Matters: Diagrammatic Renderings of Human Evolution and Diversity in Physical, Serological and Molecular Anthropology." *History of the Human Sciences* 28 (5): 3–16.

Lynch, Michael, and Steve Woolgar. 2014. "Preface." *Representation in Scientific Practice Revisited*, edited by Catelijne Coopmans, Janet Vertesi, Michael E. Lynch and Steve Woolgar. VII–IX.

Lynch, Michael, and Steve Woolgar. 1990. "Introduction: Sociological Orientations to Representational Practice in Science." In *Representation in Scientific Practice Revisited*, edited by Catelijne Coopmans, Janet Vertesi, Michael E. Lynch and Steve Woolgar. 1–19.

Lynch, Michael. 2014. "Representation in Formation." In *Representation in Scientific Practice Revisited*, edited by Catelijne Coopmans, Janet Vertesi, Michael E. Lynch and Steve Woolgar. 323–329.

Lynch, Michael. 1985. "Discipline and the Material Form of Images: An Analysis of Scientific Visibility." *Social Studies of Science* 15 (1): 37–66.

Lynch, Michael. 1988. "The Externalized Retina: Selection and Mathematization in the Visual Documentation of Objects in the Life Sciences." *Human Studies* 11: 201–234.

Martínez-Cruz, Begoña, Isabel Mendizabal, Christine Harmant et al. 2016. "Origins, Admixture and Founder Lineages in European Roma." *European Journal of Human Genetics* 24 (6): 937–943.

Mašindová, Ivica, Andrea Šoltýsová, Lukáš Varga et al. 2015. "MARVELD2 (DFNB49) Mutations in the Hearing Impaired Central European Roma Population – Prevalence, Clinical Impact and the Common Origin." *PloS one* 10 (4): e0124232.

M'charek, Amade. 2022. "Race and Sameness: on the Limits of Beyond Race and the Art of Staying with the Trouble." *Comparative Migration Studies* 10 (1): 1–16.

M'charek Amade. 2013. "Beyond Fact or Fiction: On the Materiality of Race in Practice." *Cultural Anthropology* 28 (3): 420–442.

Melegh, Bela I., Zsolt Banfai, Kinga Hadzsiev, and Attila Miseta. 2017. "Refining the South Asian Origin of the Romani people." *BMC Genetics* 18 (1): 1–13.

Mendizabal, Isabel, Oscar Lao, Urko M. Marigorta et al. 2012. "Reconstructing the Population History of European Romani from Genome-wide Data." *Current Biology* 22 (24): 2342–2349.

Milanov, Ivan, Nikolay Topalov, and T. S. Kmetski. 1999. "Prevalence of Multiple Sclerosis in Gypsies and Bulgarians." *Neuroepidemiology* 18 (4): 218–222.

Moorjani, Priya, Nick Patterson, Po-Ru Loh et al. 2013. "Reconstructing Roma History from Genome-wide data." *PloS one* 8 (3): e58633.

Morar, Bharti, Sashka Zhelyazkova, Dimitar N Azmanov et al. 2011. "A Novel GEFS+ Locus on 12p13. 33 in a Large Roma Family." *Epilepsy Research* 97: 198–207.

Parson, Walther, and Lutz Roewer. 2010 "Publication of Population Data of Linearly Inherited DNA Markers in the International Journal of Legal Medicine." *International Journal of Legal Medicine* 124 (5): 505–9.

Pickering, Andrew. 1995. *The Mangle of Practice. Time, Agency and Science*. Chicago: University of Chicago Press.

Ramal, L. M., M. A. López-Nevot, J. M. Sabio et al. 2004. "Systemic Lupus Erythematosus in Southern Spain: a Comparative Clinical and Genetic Study between Caucasian and Gypsy Patients." *Lupus* 13 (12): 934–940.

Roewer, Lutz, M. Krawczak, S. Willuweit et al. 2001. "Online Reference Database of European Y-chromosomal Short Tandem Repeat (STR) Haplotypes." *Forensic Science International* 118 (2–3): 106–113.

Ruan, J., O. Peyruchaud, P. Nurden et al. (1998). "Family Screening for a Glanzmann's Mutation Using PCR-SSCP Thrombasthenia." *Platelets* 9 (2): 129–136.

Schlegel, Nicole, O. Gayet, M. C. Morel-Kopp et al. 1995. "The Molecular Genetic Basis of Glanzmann's Thrombasthenia in a Gypsy Population in France: Identification of a New Mutation on the alpha IIb Gene." *Blood* 86 (3): 977–982.

Sekula, Allan. 1984. *Photography Against the Grain: Essays and Photo Works 1973–1983*. Halifax: The Press of Nova Scotia College of Art and Design.

Sirugo, Giorgio, Scott M. Williams, and Sarah A. Tishkoff. 2019. "The Missing Diversity in Human Genetic Studies." *Cell* 177 (1): 26–31.

Sommer, Marianne. 2015. "Population-genetic Trees, Maps, and Narratives of the great Human Diasporas." *History of the Human Sciences* 28 (5): 108–145.

Sontag, Susan. 1977. *On Photography.* New York: Farrar, Straus and Giroux.

Surdu, Mihai. 2019. "Why the 'Real' Numbers on Roma Are Fictitious: Revisiting Practices of Ethnic Quantification." *Ethnicities* 19 (3): 486–502.

Surdu, Mihai. 2016. *Those Who Count. Expert Practices of Roma Classification.* Budapest: Central European University Press.

Surdu, Mihai, and Martin Kovats. 2015. "Roma Identity as an Expert-Political Construction." *Social Inclusion* 3 (5): 5.

Star, Susan Leigh. 1983. "Simplification in Scientific Work: An Example from Neuroscience Research." *Social Studies of Science* 13 (2): 205–228.

Tufte, Edward R. 2001. *The Visual Display of Quantitative Information*. Cheshire, CT: Graphics Press.

Werner, Gabriele. 2015. 'Discourses about Pictures: Considerations on the Particular Challenges Natural-Scientific Pictures Pose for the Theory of the Picture." In *The Technical Image*, edited by Horst Bredekamp, Vera Dunkel, and Birgit Schneider. 8–13.

Woolgar, Steve. 2014. "Struggles with Representation: Could It Be Otherwise?" In *Representation in Scientific Practice Revisited*, edited by Catelijne Coopmans, Janet Vertesi, Michael E. Lynch and Steve Woolgar. 329–333.