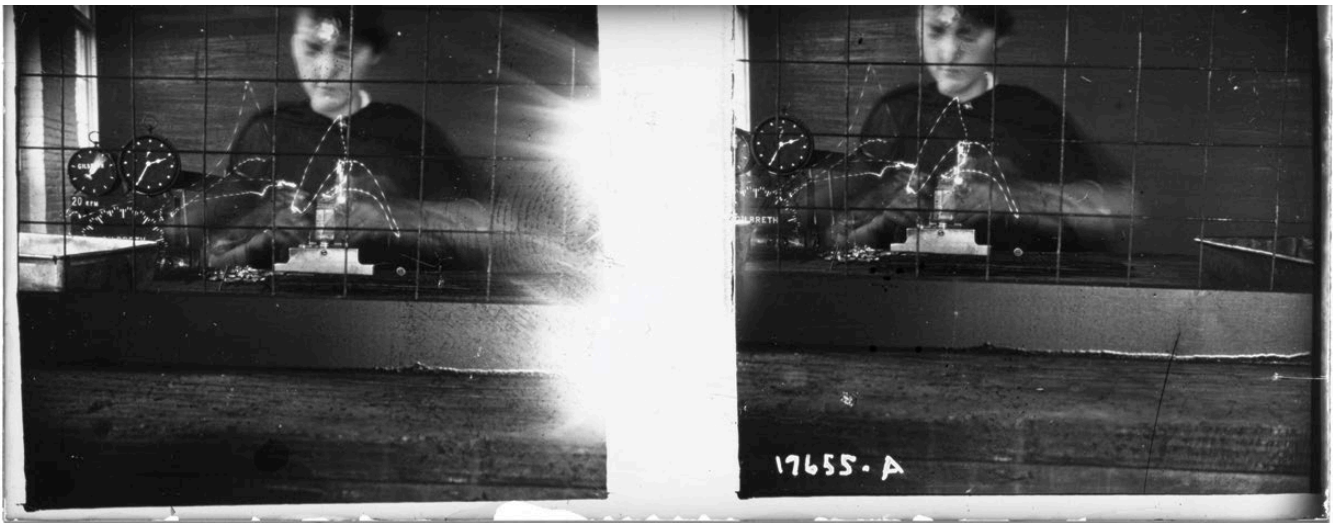


MONA SLOANE, ABIGAIL Z. JACOBS, AND EMANUEL MOSS

# The Measured Body

Redesigning motion capture systems to be more representative of real human bodies and movements could make them fairer and more useful for applications including law enforcement and medical diagnostics.



FRANK AND LILLIAN GILBRETH, *Cyclegraph of woman doing light assembly work (staking buttons, 8/23/1917 ?)*  
Three motion clocks are visible in image. Image courtesy Frank and Lillian Gilbreth Collection, Archives Center,  
National Museum of American History, Smithsonian Institution.

Following World War II, the US Air Force funded two separate projects to study the movements of a human body under certain stressors in the cockpit of a fighter jet. One research team focused on pilots' range of motion in the cockpit, and the other set out to design better impact protection systems. Over the course of their work, both teams acquired and dismembered the cadavers of older white males—eight in one study; six in the other—to collect measurements of their body segments, including height, weight, limb length, and limb volume. The study groups then used these measurements to develop models inferring the force required to generate certain human motions. Despite the limitations of the sample, researchers today are still using these same models to design and test new systems of *motion capture*, or processes for recognizing, estimating, and predicting human motion and activity.

Motion capture (mocap) technology has become so ubiquitous that most people encounter it routinely without realizing. Not only does it underpin specialized applications, including animation, manufacturing safety, medical diagnostics, and injury rehabilitation support, it also is embedded in smart televisions, phones, and video conferencing systems such as FaceTime and Zoom, which can recognize and translate hand gestures into emojis. Some applications under development for both personal devices and public services use complex detection methods that can even interpret the context of motion. For example, the Magic AI Fitness Smart Mirror acts as a computer vision and AI-powered personal trainer, providing movement corrections during home workouts. The video surveillance company Sirix markets its AI video products as being capable of detecting violence in schools,

public transit, and workplaces. And other researchers are training neural networks to monitor video surveillance for indications of potentially violent behavior, simply based on pedestrian movements.

Motion capture is increasingly used to understand and anticipate the movements of a panoply of real-life human bodies. But the technology's ability depends less on direct observation and more on layers of representations of the human form. Each mocap application uses prior representations of bodies and movements and creates new ones that it uses to interpret the world. Developers of those representations make assumptions about what constitutes a human body—and whose bodies are typical or sufficiently representative. When motion capture designers conceptualize bodies and their movement as the unit of inference, they establish an assumption of “normal” that reinforces potentially rigid ideals of what a human body looks like—for instance, that all humans are bilaterally symmetrical, or that body proportions scale across heights and weights. The assumptions

### Body worldmaking

In practice, motion capture turns the complexity of human movement into useable data. Many modern mocap systems rely on the placement of reflective markers at predetermined places on a moving body, which are then tracked by an array of cameras and mapped onto a digital skeletal model. Then those measurements are combined with parameters for aspects of the body such as limb length, weight distribution, and joint flexibility that are derived from previous models (like the aforementioned cadaver measurements) to inform a complete representation of a human body.

Motion capture systems are increasingly developed and used in public contexts. But the models these systems rely on were designed to suggest norms for body shape and mobility—not to represent a wide range of bodies and their real-world interactions. In addition to the two Air Force–supported studies, we examined three other canonical datasets for training and evaluation of motion capture tasks. Each contained revealing limitations.

## When motion capture designers conceptualize bodies and their movement as the unit of inference, they establish an assumption of “normal” that reinforces potentially rigid ideals of what a human body looks like.

underpinning most motion capture systems have been underexamined, despite their importance in shaping the human body within our collective sociotechnical imagination.

As motion capture expands into the public sphere and is adopted for entertainment, law enforcement, employment, safety, and other uses, these assumptions require scrutiny—and change. As researchers in the fields of sociology, information science, and anthropology studying mocap technologies, we see a clear need to redesign motion capture systems in the public interest. Doing this will not be easy, in part because the technology built on old models is rapidly maturing. But that is precisely why it should be done: Mocap models and the representations used to build them have extraordinarily long lives because they create and validate new systems.

Another reason to undertake this redesign is that mocap is just one of several data-centric processes that requires a reorientation from private to public interests, in which ethics and other forms of accountability can play a more effective role. To address issues buried deep in the mocap models, we recommend new ways of gathering data and involving communities in order to realign the technology with public interests—and this experience could serve as a model for reforming similar applications.

A dataset released in 2014 called Human3.6M draws on data from a sample of only 11 subjects. The subjects were actors (six male, five female) recruited to enact 17 predefined “scenarios,” such as “eating,” “drinking,” “walking dog,” and “taking photos.” (In the latter two scenarios, the actors pantomime the presence of a dog or camera.) The researchers responsible for the dataset assume the sample yields “a moderate amount of body shape variability as well as different ranges of mobility,” but the small sample size and the fact that the actors were not engaging in real-world situations limits the applicability of the sample. The assumption that the entire range of human differences in body shape and mobility can be adequately represented by 11 individuals elides the experiences of people who move differently than systems are trained to expect.

The popular Carnegie Mellon University Graphics Lab Motion Capture Database uses a larger sample size: 144 subjects. However, the subjects in this case correspond not to unique individuals, but to combinations of movements—such as modern dance, recreation, and pantomiming animal behaviors—performed between 1 and 68 times each by different people in a lab setting. In many cases, the same person performed the movement in different sessions, but no demographic information on

the individuals that participated in the study is provided. This approach also conflates imitation movements observed in a lab with genuine human movement in situ.

Another frequently used dataset originated from the 2002 Civilian American and European Surface Anthropometry Resource (CAESAR) project, which created 3D scans of a sample of civilians from three countries (the United States, the Netherlands, and Italy) to extrapolate population information for all NATO countries. The sample selection design is articulated in the survey's report: "The United States was chosen because it has the largest and the most diverse population in NATO. The Netherlands was chosen because it has the tallest population in NATO, and Italy was chosen because it has one of the shortest populations in NATO." This makes clear the strong assumptions about diversity and representation the CAESAR team hoped to capture: the top and bottom height ranges in Europe and North America.

As mocap expands into more domains, the same datasets and their inherited inferences about body measurements are pushed into new work. Thus, each new approach is

for workplace safety might receive too few (or too many) warnings about how they bend and lift, affecting their job status or changing their behavior to suit the software. A "normative" model may fail to capture a wheelchair or account for the movements of a disabled body, a cyborg body, a pregnant body, or an above- or below-average sized body, effectively erasing their presence. Given the number of proposed mocap applications for pose and gesture recognition using artificial intelligence, including for projects to support human-machine collaboration and human-centric digital twins, the list of potential harms and limitations is likely to grow.

The social assumptions baked into technology—and its design and testing—always have consequences for users. These consequences show up across domains: blood oxygen sensors giving unreliable measurements on darker skin, for example, and seatbelts built for the average adult male crash test dummy. The problem is that excavating the assumptions underlying existing datasets and systems requires careful analysis of many sources ranging from vague or imprecise marketing language to dense academic papers' methods

## The models these systems rely on were designed to suggest norms for body shape and mobility—not to represent a wide range of bodies and their real-world interactions.

developmentally linked to, trained on, and validated by earlier motion capture technologies. And even newer 3D datasets rarely include more representative populations, while others reuse previous data in various composite or synthetic datasets. For example, the popular Synthetic hUmans foR REAL tasks, or SURREAL, dataset generated synthetic 3D and 2D data from the Carnegie Mellon database and the CAESAR survey by extrapolating from limited samples to produce greater volumes of data, without broadening the range of body shapes and movements included. Even when datasets do include a variety of body measurements, they are nevertheless validated on "gold standard" measurements that still rely heavily on the original body parameters of the white, male cadavers. Thus, motion capture technologies are, by design, overgeneralizing about the typicality of human bodies and their motion.

### The social life of assumptions

Neglecting to design mocap systems for all bodies may reduce costs, but as a design choice it's a high-stakes gamble with public trust. In medical diagnostics, for example, accurate capture and analysis of body data can mean the difference between rehabilitation and further injury. Warehouse workers whose movements are closely monitored

sections. This kind of analysis requires time and skill, and mocap developers are rarely, if ever, trained to uncover such buried assumptions, let alone interrogate how their own assumptions shape their projects.

To adjust the assumptions within motion capture systems, an effort beyond addressing so-called tech ethics or AI accountability is necessary. Such an undertaking will require new assessment approaches that allow the broader research community to examine, audit, challenge, and mitigate assumptions shaping the technology's ability to create representations of the human body.

### Mocap for the people

One way to develop a new and more transparent epistemology for mocap—and to correct for the normative assumptions that have crept into mocap innovation—would be to commit to a radically different method for collecting data on what all kinds of living, breathing, moving bodies look like across the United States. A diverse and ever-growing dataset could redefine motion capture systems' inferred parameters with more accuracy for a much broader segment of the population.

Imagine a truck pulling up to a town square somewhere in the rural United States. The truck is big and brightly

colored, attracting attention wherever it goes. Inside is a mobile motion capture system, akin to mobile mammography units deployed in Europe and parts of the United States as public health services for rural or underserved communities. Through partnerships with local schools, employers, and community organizations, the arrival of the truck is announced ahead of time. The truck hosts a small exhibit that explains the history of motion capture and invites viewers to become participants. These volunteers then fill out forms indicating how their motion capture data can be used—for athletics or animation or surveillance. In exchange for their data donation, participants receive a recording of their motion capture, as well as a short animation with a character of their choice. A fleet of mocap trucks could reach a wide range of communities.

We imagine this mocap truck as a literal vehicle for public participation in the creation of new sociotechnical data, a way to collect a vast array of body measurements into a single dataset.

Compiling a large volunteer dataset of this sort brings with it great responsibility to prevent harms. It could be seen

circumstances to ensure that data is fit for purpose. These best practices consider how to document the populations represented in the dataset as well as how to manage data collection so that the data can act as a meaningful basis of comparison and validation for a wide range of motion capture applications.

But additional governance is still necessary, not only to protect the privacy of those represented in the dataset, but also to ensure the dataset is used in the public interest. Mechanisms for improving algorithmic accountability, like embedding public interest provisions into licensing and procurement agreements and designing governing bodies to conduct ethical impact assessments, are promising approaches that could be extended to motion capture systems. It is worth considering how representative, democratic entities other than federal agencies might be in a position to equitably manage such a vast mocap dataset. Local governments—closely tied to constituents and responsive to stakeholder input—could be important partners for such an initiative.

Structured with privacy and public interest protections and built with respect for people's autonomy, a comprehensive

## A diverse and ever-growing dataset could redefine motion capture systems' inferred parameters with more accuracy for a much broader segment of the population.

as a “honeypot” vulnerable to theft or misappropriation, or it could be used to build products that actively endanger or seek to identify those who contribute to it. Large datasets containing in-depth information about human difference have also been used to reinvigorate harmful claims about the biological basis of race, even while giving individuals much-desired insights into their own biology and ancestry. Data collection projects for the development of large language models are another example of initiatives that have not always been designed to benefit participating communities.

But bringing the tools of knowledge production closer to people's lives and letting them co-determine the conditions for the tools' deployment is an approach to research with some precedent. Researchers developing sign language recognition technologies have worked closely with deaf communities to collect the data needed to train such tools by recording demonstrations of individuals' signs in ways that ensure balanced representation, adequately informed consent, appropriate levels of financial compensation, and—crucially—the ability to review, edit, and delete their contributions to the dataset.

Researchers at the National Institute of Standards and Technology have developed guidelines for collecting biometric data and managing risks under controlled

dataset of human body measurements and movement difference could make mocap applications safer and more effective. Companies that develop hardware and software stand to benefit the most from widespread deployments of a range of mocap applications and would be well-served by investing in the effort needed to make deployments safer and more effective for a broader range of people.

A successful initiative to simultaneously inform and engage the public in the process of creating a dataset could also serve as a model for other data-centric systems, such as audio recordings used to train speech-to-text transcription software or education data used to predict student performance. And as data from the project enters mocap applications, it would bring into being a new representation of the human body, in all its variety, so that tomorrow's technologies are both from people and for people.

***Mona Sloane** is an assistant professor of data science and media studies at the University of Virginia. **Abigail Z. Jacobs** is assistant professor of information in the School of Information and an assistant professor of complex systems in the College of Literature, Science, and the Arts at the University of Michigan. **Emanuel Moss** is a research scientist at Intel Labs and adjunct professor of data science at the University of Virginia.*