

Introducing the Kavli HUMAN Project

April 27, 2016



A massively ambitious new research collaboration may soon become psychological science's answer to the Human Genome Project. The goal of the [Kavli HUMAN Project](#), a new collaboration between New York University's (NYU) Institute for the Interdisciplinary Study of Decision Making (IISDM) and the Kavli Foundation, is no less than to quantify every major biobehavioral factor that plays a role in shaping humanity. By comprehensively studying a cohort of 10,000 New Yorkers over the course of 20 years, the project hopes to expand the scope of social science research capabilities — just as the Human Genome Project revolutionized the field of genetics.

New technology is now allowing scientists to collect data on a scale previously unimaginable: The Kavli HUMAN Project, under the direction of APS Fellow [Paul W. Glimcher](#), wants to harness this capability to allow researchers to quantify the complex interplay of factors and feedback mechanisms between our behavior, biology, and the environment. Collecting a vast range of precise and detailed measurements on people's day-to-day behavior, environment, and health over 2 decades will create the most comprehensive picture of the human experience ever captured — from social networks and cognitive aging to genetics and the microbiome.

“Until now, no research team has ever tried to put together a single study on all there is to know about a large group of people. Kavli HUMAN Project is that effort,” [said Glimcher](#), an NYU professor of neural science, economics, and psychology.

One of the goals of this project is to not only make the most comprehensive database of human behavior ever collected, but to also make it easily available to the research community. The Kavli HUMAN Project will be the first longitudinal biobehavioral study that will provide a unique data platform assuring that researchers across all disciplines can analyze the data.

As a long-term objective, the Kavli HUMAN Project hopes to tackle — and potentially solve — some of the most challenging, fundamental questions about what it means to be human. Not only will this vast

dataset illuminate the psychosocial, biological, and genetic factors that cause us to be healthy or sick, it will be able to inform evidence-based policy decisions that can intelligently serve the public.

“Such a longitudinal data set would, for example, definitively identify the many interlocking causes of our obesity epidemic,” according to the [preliminary study design](#). “It would finally allow us to determine the critical factors in successful childhood development. It would define when and why people suffer from Alzheimer’s disease by uniting data about environmental exposure to pathogens, diet, behavior, medical care, and genetics in a single longitudinal database.”

The project will start out by recruiting 2,500 “seed” individuals from across the entire city, beginning in 2017. As the study progresses, family members from each of these initial “seed” participant households will be recruited as participating family units until the total study population of 10,000 individuals is reached.

New York City was chosen as a location in part because of the immense sophistication of the city’s data infrastructure. The project will have permanent access to the ongoing database at the Center for Urban Science and Progress, which provides detailed information on municipal data ranging from transportation and energy use to law enforcement and the school system.

Once enrolled in the study, basic biological and physiological data will be collected through medical exams. Throughout enrollment in the study, a huge variety of behavioral and biological data can be collected automatically, including drug and chemical exposure, diet assessment, sleep behavior, personality traits, IQ, mental health status, long-term and working memory, employment data, and detailed financial transaction records.

Because much of the information being collected is sensitive in nature, the project has invested extensive resources towards designing a database that can provide the highest possible privacy and protection for subjects while also allowing scholars to access data for research. To meet these goals, the database has been uniquely designed to allow researchers easy access to anonymized subsets of data, stripped of all personally identifying information that could threaten participant privacy.