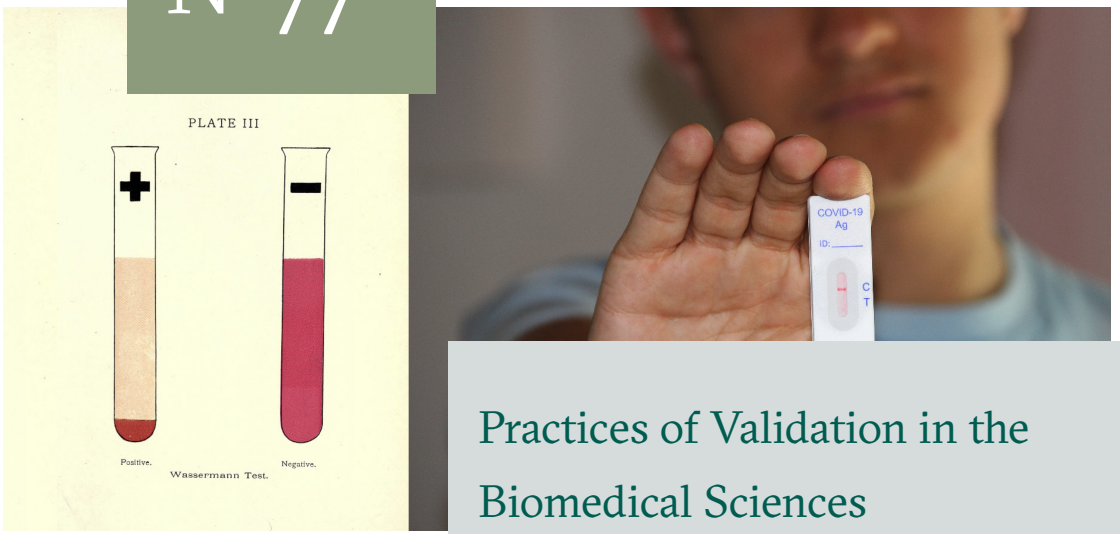
Placing History and Philosophy of Biomedical Knowledge into Perspective

The Wassermann test mentioned earlier was influential in forming biomedical understandings of laboratory-based diagnosis, but it also played an exemplary role in the early days of the history, philosophy, and sociology of science as a discipline. It formed the topic of Ludwik Fleck's 1935 book *Entstehung und Entwicklung einer wissenschaftlichen Tatsache,* which was later praised as inspiration for the prominent account of "paradigm shifts" (Kuhn) in the history of science. Fleck was a trained physician and serologist. We take this as an occasion to reflect upon the intersection of "actor" (e.g., biomedical researchers) and "analyst" (e.g., philosophers of science) perspectives in the historical genesis of validity concepts. For instance, Paul Meehl (1920–2003) practiced as a clinical psychologist while assisting the establishment of the Minnesota Center for Philosophy of Science in the early 1950s. Meehl played a role in introducing "construct validity" as a term to signal doubt—very much in the sense of the popular philosophy of science of Karl Popper. The term was designed to help capture concerns regarding whether a test is informative about what it intends to be about.

The pursuit of validity in biomedicine has been and remains a transdisciplinary project that encompasses within it the history and philosophy of biomedical knowledge. By reflecting upon this project and our place within it, this Research Group aims at a better understanding of the history and philosophy of evaluative categories and methods—both in biomedical research and within the history of our own field.

RESEARCH TOPICS

N°77



Practices of Validation in the Biomedical Sciences

by Lara Keuck

OCTOBER 2021

In 2021, the validity of Covid-19 tests became an issue of political and public concern. Can we rely on tests to re-open schools, shops, and museums while the pandemic is still in full swing? Seeking to communicate to the public, scientific experts have explained in podcasts and newspapers the meaning of technical terms connected to a test's "validity," such as "specificity" and "sensitivity," which indicate the proportion of infected and non-infected people, respectively, that will be correctly identified as such by the test. The idea that these measures are largely context independent properties of a test, is closely related to the history of the introduction of these terms: In the early twentieth century, serologists (who study blood and other bodily fluids) articulated concepts of specificity and sensitivity as part of debates over the correct way to conduct and interpret the Wassermann test. This blood test was used as a diagnostic test against syphilis. The test's interpretation as "specific" and "sensitive" depended upon contentious assumptions about the nature of the reaction, and the different stages of the infectious disease.

ABOUT THE AUTHOR

Lara Keuck is Max Planck Research Group Leader of "Practices of Validation in the Biomedical Sciences" at the MPIWG. She holds a graduate degree in Molecular Biomedicine, a PhD in History, Philosophy, and Ethics of Medicine, was a Branco Weiss Fellow of ETH Zurich, and is a member of the Junge Akademie. Her previous research focused on the epistemology of medical classification systems, vagueness in psychiatry, and the history and philosophy of biomedical models of mental illness.

Front page: A graphical representation of the Wassermann test to detect serological reactants associated with syphilis (Maximilian Herzog, 1910: *A Text-Book on Disease-Producing Microörganisms* [sic]. Philadelphia and New York: Lea & Febiger), and a person holding a negative Covid-19 antigen test (Pixabay/Alexandra Koch 2021).

More research topics are available at:
www.mpiwg-berlin.mpg.de/researchtopics

Max Planck Institute for the History of Science
Boltzmannstraße 22 · 14195 Berlin
T +49 30 22 667 0

Later the terms were emancipated from these disease-specific assumptions and methods, and were applied in different contexts as statistical measures. Today the question of the conditions under which a test's specificity and sensitivity are valid is once more gaining attention.

The example of the specificity and sensitivity of diagnostic tests—which is also the point of departure of one of the projects in the Research Group—showcases how the problem of interpreting the meaning of test results has a long history, which has over time been sidelined, or forgotten, and reactivated. The new Max Planck Research Group "Practices of Validation in the Biomedical Sciences" examines this history. How has validity been practiced? And how has uncertainty been assessed, regulated, and argued about? We address these questions to study the development of the modern biomedical sciences, and to situate contemporary challenges of translating and evaluating biomedical knowledge.

Historicizing Validity

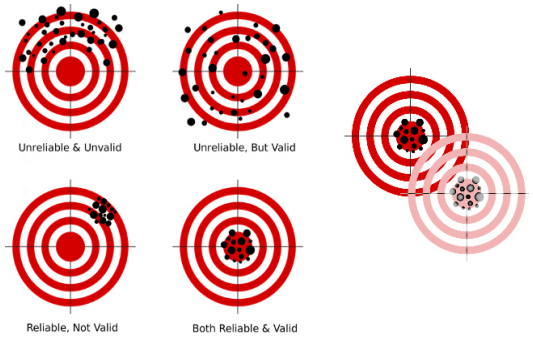
A key term of the Research Group is "validity." In the twentieth century, the technical term "validity"—one with a complex genealogy from logic and statistics to psychology—was put to use in many of the sciences. Often represented as hitting a target mark, the term was used to denote the extent to which an assessment of an item actually (but not necessarily reliably) captures its intended (abstract) target that the test is applied for. For example, psychometricians would deem an intelligence test "valid" if it was informative about the hypothetical entity of "intelligence."

This Research Group will examine the philosophical and historical foundations of this understanding of validity, as well as reconstruct the tracks on which concepts of validity and methods of validation made their way into mid-twentieth century biomedicine. Through examining how validity was practiced in the long twentieth century, we seek to understand the challenges to ascertaining how informative biomedical studies were about a medical target of interest.

For instance, we are studying how the exponential growth of research communities, the emergence of international research organizations, and technological developments, such as those associated with the Human Genome Project, impacted on the implementation of standardized validation procedures in the postwar period.

N°77

Most recently, the failure to replicate many published research results, the so-called reproducibility crisis, has been discussed as an epistemic and economic problem for the successful translation of hopeful pharmaceuticals from animal experiments to clinical trials. We take such case studies to examine how the scaling-up of biomedical research impacted discourses and methods concerned with validity in several ways.



01 Validity and reliability (Nevit Dilman, 2012/CC), and a modified version of it that represents the challenge of the moving target that the Research Group has a particular interest in.

In the past century, not only practices of validation but also biomedical approaches to human health and disease have undergone fundamental reconfigurations. We are therefore particularly interested in the history and philosophy of scientific methods that were used to ascertain the validity of research on *moving* targets, such as in psychiatric research.

First Research Projects and Working Groups

The Research Group combines philosophical and historical perspectives on the establishment and change of evaluative methods and categories. The first projects of individual group members focus on the establishment and transformations of "specificity" and "sensitivity" in diagnostic tests throughout the long twentieth century (Nicholas Binney), and examine the mid-century search for universal validation and changing scales of validity in the postwar period (Alfred Freeborn). Our sources include not only publications and archival materials, but also oral histories. With regards to the latter, we will work with a combination of digital humanities methods and qualitative historical research to explore new possibilities to re-use and -analyze already available oral history repositories. Furthermore, we are coordinating two interdisciplinary working groups, "Translating Validity in Psychiatric Research" (with Steeves Demazeux, Université Bordeaux Montaigne) and "Validation and Regulation in the Sciences of Health" (with Angela Creager, Princeton University), in which we collaboratively and comparatively examine how "validity" and "validation" have been globally applied and locally adapted to evaluate and regulate a broad range of objects from toxicological tests to psychiatric constructs.