Sociology as a population science, J. Goldthorpe

Alessandro Caggia

June 2025

Abstract

Culture is extraordinarily particular and diverse across time/-place. Expressed **through language and symbolic communication** in a variety of **values, myths, religions, rituals, art**. Central question: **how to accommodate such variability**?

1. Theories of sociological thinking

Question 1

1. Structural-functionalism / holism (macro)

+ Theory:

1. **Typological thinking** (physics, chemistry): Sociocultural entities (hence Society) can be represented as **homogenous entities** (e.g., electrons) representing **fixed wholes**, **ideal types** that, as **'things in themselves'**, **are to be taken as the key units of analysis without the need of reduction** to the individual level;
2. **Variability** treated as occurring **among sociocultural entities** (tribes, communities, classes, nations).
3. **Searches universal deterministic laws applying equally to all of them** (Hooke's law). **Variation** considered **erratic**.
4. **Functionalism** represents the main theoretical resource of the holistic paradigm:
   - **Individuals merely carry out the needs (imperatives) of the social system**, an **epiphenomenal socioculturally programmed behaviour**. **Society is superordinated with respect to individuals, who basically are puppets in its hands. Behaviour is an expression of the system, not as something that could change or influence it**.
5. **Methodological holism** (Causation): Such **Social fact** can be explained only **by** reference to other **social facts**.
   - Society seen as an integrated **whole processing through ENDOGENOUS PREDETERMINED STAGES OF EVOLUTION** (**from feudal to capitalist to socialist; from premodern to modern etc**). Example: **Marxist theory** of capitalist crises and revolutions.
     - Such logic does not depend on individuals: ex. **you cannot do a revolution until capitalism** in that country has not reached the stage of development enabling a revolution. Indeed they were **against Lenin**: Russia was not yet ready.
     - Example: the creation and expansion of modern school systems, **Modern schools as a function of industrial development**: manufacture develops, skills are needed in order to produce → schools created.
     - Very far from our view: we explain macro phoenomena in terms of individual action.
6. **Institutionalist perspective**: Institutions are not only given to and **independent** from individuals (are social facts) but also **produce** the individual (through norms, education etc). Instiutions **make social order possible** and are way **more than the summatory** of the parts (as in textbook micro). Example: Italy exists independently from Italians. Basically institutions **have their own ontology**. Very far from our view: we usually adopt a **nominalist** perspective, rooted in **symbolic interactionism** and **Kantian epistemology**: we see entities as Italy as an abstraction from Italians, with Italy existing as a way our mind work to make its job simpler when considering Italians. Overall question:
   - What exists first the individual or society (i.e. what is the ontological root)? Italians or Italy?

Remark: Institutions are a form of 'social facts'.

- *Social facts **are ways of acting, thinking and feeling which possess the remarkable property of existing outside the consciousness of the individual**.*

Within the framework discussed, Durkheim conceived **Sociology as the Science of Institutions**. **Key strengths:** the recognition of **unavoidable external constraints** on individual behavior.

- Critical Points:

1. **Theoretically: Neglect of individual variability** *within* entities by assuming high internal homogeneity (consensus, conformity). **Changes are seen as endogenous to a macro actor and not as result of actions of individuals.** Example: Marxism: individual entrepreneurs are just like puppets acted on by capitalism, the super subject. LENIN COUNTEREXAMPLE. **Theoretical basis** (Parsons, 1952) shared beliefs and values are **institutionalised** in social structure while at the same time they are **internalised** through enculturation. Thus knowledge of institutional forms can in itself provide an adequate enough synopsis of prevailing patterns of social action.
2. Empirically: Studies of **religious and political beliefs**: Even with subcultural and group affiliations included, only a modest part of variation is explained. Example (**Malinowski**):
   - Holism claims that in primitive societies the individual is **completely dominated by the group**, that he obeys the commands of his community, its traditions, its public opinion, its decrees with a slavish, fascinated, passive obedience.
   - **Trobrianders**: truly, **sociates are divided, unstable, changing internally and externally**. Tribes cannot be seen as perennial/ in an indeterminate age.

= Conclusion: Holistic paradigm tried to **endogenise ends of action** (show such ends of action are derived by social groups, culture and norms, ie endogenous); but failed, individuals ends and values cannot be fully reduced to group-level determinants. **Better accept exogeneity** of ends/preferences as in mainstream economics; **Few** still believe possible to formulate deterministic laws. Regularities are diverse and complex; micro, meso, macro context dependence (as represented, say, by social networks, associations, or by institutional and wider social structure), differences across subpopulations (age, gender, ethnicity, occupation).

Question 2

Conflict theories (macro)

- **Opposed to structural-functionalism: regards social order as unstable, society divided into groups endowed with different levels of resources who compete**.
- **Key strength: realistic** picture, rooted in the experience of war & revolution.
- **Key weakness:** often reduced to a **single dimension** of conflict.
  - In reality, social conflict and social stratification are **multi-dimensional** (many types of resources). Many forms of competition: recognition, social status, wealth, identity, etc.

Question 3

Symbolic interactionism (micro)

- **Society** and the world are a **subjective creation**, or betterm a result of a **social process**.
- Derives from:
  1. **Radical empiricism: Cartesio** cogito ergo sum.
  2. **Thomas theorem**: "If men define situations as real, they are real in their consequences"
- **Key strength: key to explain collective identities**, the societal relevance of symbols and rituals. Focus on **subjectivity** and **inter-subjectivity** (bc we create things).
- **Key weakness:** possible identification between *subjective*, *volatizable*, and *voluntary*.
  - Kants: "subjective" does not mean "voluntary" or "volatile"
  - the world as a social construction is stable.

2. The individualistic Paradigm (Utilitarism / Rational Choice theory, RTC), micro

Question 4

How does methodological individualism account for individuality?

- **Micro-level theory**, based on **individual choice** (and marginal utility theory). Society is the result of rational actions guided by **maximization of utility**.
- **Key strength: flexibility, consistency** with lived experience (everyday we see people trying to maximize their utility). PROVIDES MICROFUNDATIONS

Expand utility

- Include **collective utility maximization** (e.g., marriage).
- **Expand utility**: Material needs are linked to physical needs. Individuals compete not only for material rewards but also for **immaterial benefits**: **identity**, **status (prestige, reputation, honor)**, power.

Expand constraints

- **Internal constraints**:
  - **Behavioral Beliefs**: attitude towards the behavior.
  - **Normative Beliefs**: social pressure from others (perceived).
  - **Perceived Behavioral Control**: perceived ease of action.
- **External constraints**:
  - **Actual behavioral control** (police, law).
  - **Normative constraints: Social norms & structure**: shape both utility and constraints; type of sanction matters.
  - **Non-normative constraints: factual constrains, Resources**, eg housing costs are a c constrain for cohabitation.
- **Relax rationality** (see below).
- **Key weaknesses:**
  - Assume **rationality** to simplify math
  - **Difficult to integrate different dimensions of utility** (e.g. identity, status). Measurement of immaterial utility is difficult.

Question 5

The individualistic Paradigm

A crucial distinctive feature of humans is **individuality**. **Behaviour** of individuals **cannot be simply derived from the knowledge of institutional** forms. They have to be studied through specific methods that:

- i. **Allow** to assess the full extent of **variability** that exists within entities
- ii. **Allow** to derive probabilistic **regularities** from this variability

COLEMAN COMPLIANT

Question 6

The roots of Individuality

- **The theoretical foundations** of the individualistic paradigm and **the generative force** of the sociocultural phenomena lie in two ingredients: 1) humans have **distinct personal ends**, 2) **Autonomy in human action**.
- **Informed choice** (mental ability to *prefigure* actions, assess situations, and evaluate future consequences), coupled with autonomy in action, is the **instrument that makes the realization** of distinct personal ends possible **through** trade off, cost benefit, **evaluation** of, among which, norm deviant or compliant behaviour. **Such informed action is rational** ina way different from economic hyper-rationality (demonic):
  - **No unlimited information/calculation power**. Rationality has costs. Collection and computation are costly (the brain requires a lot of energy). No *Homo Economicus*, rather *homo sapiens sapiens*.
  - **Bounded rationality**: Humans seldom have access to all relevant information. Focus on satisfactory outcomes under uncertainty. Satisfy + suffice = **satisficing rather than optimising** (Herbert Simon), **one sets an aspiration level and stops when one has found a choice above this level**.
  - **Psychological standpoint**: Focus on **procedural** aspects, how people actually make decisions. Key idea: people resort to an **adaptive toolbox filled with fast (little time) and frugal (little info) heuristics (simple rule of thumb).** Example in medicine: is blood pressure > 91?
    - No: ok no further exam
    - Yes: is age > 62.5?
    - No: ok, chill
    - Yes: is there also tachycardia?

    Heuristics first seen (Tversky & Kahneman 1974) as sources of **error/irrationality**. **Adaptive toolbox view: heuristics are specialized, efficient tools shaped by evolution and learning.** Even though they are simple, these heuristics often produce **good outcomes in stable environments**, performing as well as and even better than traditional rational modls of optimization. But sometimes can have **bad collective consequences** (e.g., **panic** behavior: financial / at theater etc) and **do not work in environments changing fast**.
  - **Sociological standpoint**: focus on **situational** aspect. Individuals can be understood as **acting appropriately for the attainment of their ends** considering their informational or time constraints. Two modes:

- **Thinking slow**: considering all aspects, raising information, processing it carefully.
- **Thinking fast (children)**: relying on a "portfolio" of traits: men+ race + foreigner. This is the root of the persistence of stereotypes and of a number of 'irrational' choices.

– Remark: **informed, subjectively rational choice may itself often lead to conformity** with established norms and practices (eg Do what the others do), as it is normal for norms to be accepted and followed as they solve coordination problems, free riding and allow for provision of public goods.

Question 7

What is the difference between Methodological and Ontological Individualism? Boudon (1990)

– **No ontological claim**: **does not claim that Individuals only exist**, no denial of the reality of sociocultural phenomena (that "there is not such thing as society").
– **Methodological claim**: **Trivially true, where else in human social life actual causal capacity could lie if not with the action of individuals**? Sociocultural phenomena must, in the last analysis, be accounted for in terms of present or past, intended or unintended, direct or indirect individual action (causal primacy of Individual Action). **A failure in doing so may arise either 1)** due to a failure to see that methodological individualism does not entail **ontological individualism or 2) because of an insistence that individual action is always influenced by the social conditions under which it occurs trivially true**, but without being in any way damaging to methodological individualism (does not require an *individualistic system of values*).

Constraints in the individualistic paradigm a

– Choices realistically open are **constrained** by inequality. Within the holistic paradigm the distinction between constraint and choice in individual action becomes blurred, if not lost. The individualistic paradigm instead can easily accomodate that social norms may be experienced as constraint and non normative constraints.
– Example
  * In holistic theories, stratification challenges the assumption of homogeneity.
  * *Fonctionnalisme rose* (Parsons): stratification = functionally necessary, normatively sanctioned allocation of the most able to key roles.
  * *Fonctionnalisme noir* (Marxist): stratification = inherent to social systems, maintained through domination of inferior by superior classes.

3. Population Regularities as Basic Explananda

Question 8

What is the difference between Events and Regularities?

- **Basic explananda: The main task of the social sciences is to explain social phenomena, that is, events that can be shown to occur within a given population with some degree of regularity**, that is, probabilistic **population regularities**.
  Examples: 1. Regularity "Why are poor people less likely to emigrate?" 2. Singular event: "Why did President Chirac call early elections in 1997, only to lose his majority?" [1]

[1] Elster also gives examples which, while referring to singular events, are, it seems, to be understood as particular instances of regularities: Kitty Genovese, bystander effect

Question 9

Chance and Explananda

- Difficulties arise when **singular events** replace regularities as explananda due to *chance in social life*.
- **Operational Chance:**
  – Chance defined in the **probabilistic** (Neymann) sense. Such probabilistic chance **can be "tamed"** (addestrata): **despite the pervasiveness of essential chance large number populations allow regularities emerge**. Such regularities are our explenanda. **Example**: We cannot predict the exact **life trajectory** of every individual aggregate mortality probabilistically.

Essential Chance:

- **Radical idea: Chance as a fundamental property of reality**. Outcomes result from **intersection of two or more independent events**. **Even if each series is determined, their independence makes the joint outcome a *"coïncidence absolue"***. his cannot be "tamed" into population-level regularities; **it's a one-off intersection**. Essential chance shapes the outcome.
- **Essential chance dominates singular events**. Examples:
  – Example: a. Dr. Dupont goes on a call. b. at the same time, roofer Dubois drops a hammer. The hammer kills Dupont. Each event sequence (doctor's movement, roofer's work) may be fully determined. But their independence makes the outcome (Dupont's death) a pure coincidence ("coincidence absolue").
  – Other examples: inherently complex and improbable intersections of preceding events (World wars; collapse of Soviet Union, Challenger disaster).
- **A priori, predict singular events is inherently improbable, although it may be possible, ex post, to suggest a narrative of specific, place and time linked factors that could have been conducive to such events. Such explanation is historical (Popper)**. Hence sociology analysis should not focus on explaining them.

Question 10

Example of Historical Explanation

- **Social scientists do not often attempt** to explain singular events, BUT still sociologists seek to explain events or complexes of events that are **grouped together under some rubric** as if they were characterised by significant regularities, but where no compelling demonstration of this has been provided. **Constructed populations: The description of population regularities and their explanation are confounded; only those cases are considered where a model can be shown to fit**, models that, like an ill-cut suit, just fit where they touch.
  They end up giving tautological, ex post, historical explanations.
- Example: **Sociology of Revolutions**. **Goldstone: progress via inductive case studies of "finite sets" of revolutions, not general samples:** *Conjunctural process model: society enters "revolutionary situation" when: (i) state loses resources/obedience; (ii) elites divided; (iii) large population proportion mobilised.*

Critique:

- **Tilly: conditions so close to defining revolutions that Goldstones model explains little.**
- Indeed Goldstone: conditions have **no inherent tendency** to converge;

4. Sociology as a Population Science

Question 11

Define the Concept of Sociology

- Sociology should be understood as a *population science* in the sense of Neyman: **Populations** = *categories of entities* satisfying certain definitions but varying in their individual properties. Examples: *human, animal, molecules, galaxies.*
- **Individual elements** are subject to considerable variability and might appear indeterminate in their states and behaviour. However, **populations** exhibit aggregate-level regularities of a probabilistic kind (so we can more easily forecast behavior of individuals). Ex. age of death etc.
  In sociology, an **individualistic paradigm is necessary** to account for the high variability observed at the individual level .

Question 12

What are the aims of Population Science?

Aims are **twofold**:

1. **Empirical (visibility)**: investigate and establish *probabilistic regularities* of populations/subpopulations **using statistical data collection and analysis**. **Statistics makes Regularities** = explananda of population science **visible**.
2. **Explanatory (transparency)**: determine **processes/mechanisms** at the individual level producing these regularities.
3. Some regularities: *highly visible and transparent*. Traffic marked regularity in car numbers 79 a.m. weekdays; decline on weekends. Shows constraints/opportunities shaping actions. More typical in sociology: less visible/transparent regularities, requiring *considerable effort in data collection/analysis*. Example: Quetelets regularities only possible thanks to national *official statistics* (censuses, registration systems).
4. one of the most discussed regularities is the demographic transition.

Question 13

Analytical approach:

- We can decompose the **structure of the methodological individualism** in two stages (**Coleman boat**):
  1. Stage 1 (macro): We **discover some regularities** at the population-level that have a certain probability distribution.
  2. Stage 2 (micro): **We explain these probabilistic regularities through individual action.**
- **Bridging the micro-macro gap is crucial to obtain sociological explanations.**

5. Statistics, Concepts and the Objects of Sociological Study

Question 14

Origins in Astronomy

- From 18th century, **astronomy** used statistics as a tool to **separate truth from error**. Say they want to **derive the orbit of jupiter: the parameters on the Newtonian equations (the "truth") + some randomic errors.**

Question 15

Statistics and sociology

- Historically, sociology as a population science is a product of the rise of **probabilistic revolution**: major intellectual shift from determinism to chance. The **erosion of determinism was followed in parallel the taming of chance**: that is, the process of making chance and its consequences intelligible and manageable.
  First Example: **Quetelet** Used Gaussian error curve/normal to display regularities in the distribution for moral statistics (marriage, illegitimacy, suicide, crime) showing that **an higher-level probabilistic order could emerge from individual actions with apparent free will/choice**.

Question 16

Statistics Creating Objects of Study

- In social sciences, statistics not just for reducing error but I. in **RCTs**, II. in **creating objects of study**.
- Quetelet: normal distribution became **substantive interest**, defining probabilistic attributes of *lhomme moyen*.

- Statistics foundational for sociology as a population science: by establishing **population regularities**, it defines the *objects of study*

Question 17

Ontological Questions

Regularities in sociology are **constructed twice**:

- 1. Through the **statistical forms** of analysis applied (Probabilities, correlations, regressions, distributions).
- 2. **Deeper**: Through the **sociological concepts** operationalized as variable used in statistical anlaysis.
  - Example: A **correlation** between education and voting exists only after concepts such as **education** and **voting** are defined.
  - **Hence, the reality of these regularities depends on conceptual decisions, making their ontological status uncertain**.

Core ontological issue: Are statistical regularities **genuine features of society** or merely **artifacts of conceptual choices**? And, are they commensurable = can we compare them?

- In **natural sciences**: existence of certain basic conceptual schemata **(natural kinds, eg classification of particles).**
- **You can believe or not that the natural world** can be carved at its joints to use Platos expression (eg think at species), **but few would regard this as being possible with the social world. Sociological concepts are (often controversial) products of human efforts to grapple cognitively with this world**.
- There is a **nominalism** flavor: **idea that statistical methods in social sciences create the objects of study rather than merely discovering them** [2].
- **Reality itself (not just categories) is socially constructed. Denies or downplays any mind-independent, objective reality. Absurd consequences**: take a body of existing scientific knowledge, eg **physics**. Extreme constructivism: **physics is not determined by how the world actually is, is contingent**, under different sociocultural circumstances **an alternative non-equivalent but no less correct physics could have developed. Nobody has eever seen it**. Unintened reduction ad absudum: Latour questioning the conclusion reached by archaeologists examining the mummy of **Ramses II** that he died, c. 1213 BC, of **tuberculosis**. Tuberculosis bacillus was only discovered that

[2]*Nominalism is a philosophical doctrine argues that concepts as "justice," are just names (*nomina*) **that humans assign to groups of things based on similarities**. However, this does **not** mean that reality is purely a social construction with no objective basis **("extreme constructivism"): the objects exist but we decide how to group and call them. So the point is that no universal concepts as independent entities exist.**

is, constructed in 1882, Latour asks if this conclusion is not **as anachronistic as claiming that Ramses death was caused by a machine gun**.
- Instead **we believe that the world has a material existence independent from our mind**, transcendental (Kant) or not. The researcher, on this reality = on empriically observable and measurable regularities, constructs concepts.
- Popper talking about "the **myth of the framework**" argues that **different conceptual frameworks** 1) **are not always incommensurable, so they can be evaluated and compared to pick the best** and 2) they can sometimes to **translate** into one another.
- Two criteria for evaluating concepts:
  - ***Applicability in research***: how far concepts can be **operationalized** through measurement instruments **with reliability and validity**.
    - **Reliability**: the degree to which a variable gives **consistent results under conditions where it should**.
    - **Validity**:
      - **Construct validity**: extent to which an instrument empirically **captures what it is conceptually supposed to capture**.
      - **Criterion validity**: extent to which an instrument **correlates with variables it should theoretically correlate with**.
  - ***Revealing power***: **how far they expose phenomena of substantive interest**.
- Despite disagreement (*the species problem*), biologists continued productive research. Best practice: avoid extreme *realism* or *nominalism*, adopt **empirically disciplined conceptual pluralism**. **It is fine to have different conceptual choices, such are judged by the results they yield. Such debate is not in abstracto but empirical**. Different measuers have different advantages and disadvantages.

Example: Social Stratification

- Many ways to investigate the issue: **income, wealth, class, education** etc.

6. Statistics and Methods of Data Collection

Question 18

Foundational Role of Statistics

- Need to **capture and accommodate variability** in individual level data. Two statistical methodologies have become central: **sample survey research** and **multivariate data analysis**.

Question 19

Emergence of the Sample Survey

- **Sample surveys are the means through which we gain the data to be analysed by our statistical models on populations**.
- Its rise solved two persistent problems in social research (mid-19th to mid-20th century):
  - Census and registration data, though invaluable, were **restricted in scope and costly**.
  - If complete enumeration had to be replaced by partial studies, new methods had to move reliably **from part to whole**.

Question 20

Early Solutions: Monographic Studies

- Advocated by **Frédéric Le Play**: **intensive case studies of families**. Emphasis on qualitative depth over number of cases. Le Play studied 36 families (later 57) across Europe;
- Aim: cases selected as **typical** of broader populations (making an appeal to the Queteletian idea). Appeals to local authorities or statistics guided selection.
- Result: mistaken generalizations (e.g. myth of the extended family as universally prevalent).

Question 21

Statistical Critiques and Advances

- Statisticians argued **typicality was mistaken in principle**: populations must reflect the **full variation of cases** one find in life (**shift from typological to population thinking, in parallel with the move from the Queteletian statistics of the average to the Galtonian statistics of variation**).
- Anders Kiaer pioneered **representative sampling**.
  - Aim: provide a **true miniature** of the target population.
  - Sampling units **chosen purposively to match** census distributions.
  - **Check representativeness against distributions of control variables** (e.g. age, marital status, occupation).
- **Limitations**
  - 1. **Bias**: Fieldworkers were required to select secondary units as those who would best represent the whole range of social variation existing within the unit.
  - 2. **Selection is not random and not quantifiable**, theres **no** mathematical basis to compute **CI** .

Question 22

From Purposive to Random Sampling

- Final solution, unbiased: shift from **purposive** to **probabilistic (random) sampling**: every individual in the target population given a known, non-zero probability of selection.
- Neyman demonstrated how prior knowledge of the target population, which had played a large role in purposive sampling, could be properly brought into sample design. He proposes **startified sampling**! (no human based selection)! [3]

Question 23

Case Studies and Sample Surveys

- **Monographs** (case studies) and **sample surveys** are two distinct methods of social data collection, each with its own **logic of moving from part to whole**.
- Case studies reflect a **holistic paradigm** as **you derive insights about the whole social system from the intensive study of one part**, while surveys reflect an **individualistic paradigm**. Why increasing dominance of survey methodology?
  - **Externalist** account: linked to macro-social changes (e.g. popular democracies, mass consumer markets.
  - However, such an externalist account apart from being **conjectural**, is seriously **deficient** in neglecting what would bean **Internalist account**: survey **were superior in solving longstanding issues, and represent a better approach as the better logic of moving from part to the whole**.**Case studies**:

[3]*systematic survey: Selects individuals at fixed intervals from a list.
Non-probability sampling is a sampling method where units are selected based on non-random criteria, meaning not every unit has a known or equal chance of being chosen. Risk of bias due to non-random selection.
*Snowball (referral) sampling: Starts with known participants, who then refer others (e.g., irregular migrants). Useful for hard-to-reach populations.

  - **1)** Issue of **internal validity**, **How to select these communities?** 2) Issue of **external validity**: how do results of my study relate to the wider population? No systematic rules exist comparable to statistical inference.
  - **1) Nice as a first explorative step**, **2)** case studies **can be generalised to theoretical propositions**, especially when chosen as *critical* or *deviant* cases.
* Only issue of sample survey: declining and biased response rates.

Question 24

Dominance of Surveys and Big Data

* **Surveys** remain crucial **benchmarks** to compare. One case is that of **Big Data**.
* Issues of big data
  - **Sample selection bias** and this reprsentativeness
  - **Not created for econ research** but business etc. **A large amount of data is in no way synonymous with a large amount of information**.
  - Great failures (Google Flu Trends Project)

Question 25

Advances in Survey Methodology

– **Improvements in survey design:**
  * *Cross-sectional surveys*: repeated surveys of the same population at different points in time.
  * *Longitudinal (panel) surveys*: follow the same individuals over time.
– **Pivotal role:** understanding processes of social change.
– Longitudinal surveys are crucial for **separating influences** on individual life-courses of time (eg cohorts eggects) and place effects.
– This distinction follows the need to separate **history** (broader context) from **biography** (individual trajectory).
– **Empirical findings:** Analyses of survey data show how different effects (period, cohort, life-cycle) generate **remarkable diversity in individuals life-courses**.
  * Supports Wrongs argument that such diversity counters the homogenising tendencies of enculturation and socialisation.
– **Hierarchical survey designs (cluster sampling):**
  * Allows comparison of:
    - *Individual-level effects* (income, education, gender).
    - *Contextual-level effects* (living in a deprived neighborhood, working in a unionized firm).
  * **Illustrative example:** A childs educational achievement depends not only on their own ability or parents education (individual factors), but also on their schools resources and peer composition (contextual factors).
    - Purpose: capture **contextual effects** on individuals life-chances and life-choices, i.e. effects of group composition and structure.
    - In case studies of holistic inspiration, it is often simply assumed that contextual effects are pervasive. Truly they are not, and contextual effects prove to be difficult to separate out from individual selection effects

Register countries

Gov monitor each transaction individuals have with the state and gains info on: schools, hospitals, income, and household tax, all linked through a unique identifier (e.g., fiscal code). Think of INVALSI.

7. Statistics and Methods of Data Analysis

Question 26

Population Regularities and Data Analysis

- You have surveys then **you have to create variables**:
  1. **Formation of appropriate concept**s.
  2. **Development of classifications/scales to operationalize** concepts.
  3. **Ensuring variables possess reliability and validity**.

Question 27

Critiques of Variable Sociology

- **First objection: In variable-based analysis, actions and interactions of individuals are lost from sight. Explanation reduces to statistical accounting**.
- **Second objection: important aspects of human social life cannot be reduced to variables**. BUT critics offer **no alternative language**.

Question 28

Variable Sociology and Mertons Requirements

- Before sociological analysis, **two requirements**:
  – **Existence** of a social regularity: events of a certain kind must have enough of a regularity to require and allow explanation, **pseudo-facts** have a way of inducing **pseudo-problems** which cannot be solved.
  – Effort must ensure **the form of regularity is properly understood**; simple regularities may hide complexity.

Question 29

Multivariate Analysis

- Used to **Demonstrate associations among variables**, initially thought to be enough to demonstrating potenital causation.
- Truly the value is primarily **descriptive**, aiming at **reliably establishing explananda**, **rather than providing causal explanations**. Making **VISIBLE for further study** regularities of a hitherto unrecognised kind.

Example: Gender Gap in British Elections

- From 19451980s: Women more Conservative than men
- Explained by a specific gender effect: Labours masculinist blinkers.
- However, true nature of regularity not clear until more detailed surveys included class and age.
- Multivariate analyses showed gender gap was **more complex** than initially observed.
- **Epiphenomenon:** Gender gap largely explained by *other regularities*, not gender-linked voting.
- Gender gap arose from:
  – Womens longer life expectancy, and especially women in more advantaged classes.
  – Older, advantaged people being more Conservative-leaning.
- **Simple bivariate associations were misleading;** multivariate analysis avoided pseudo-problems.

Digression

- Althoug regresion is the same statistical operation. Two contrasting conceptions:
  - **Blalocks Gaussian (causal/typological):** regression seeks *law-like causal relationships*; deviations are treated as noise; typological thinking assumes universal causal laws. Error is meausuremnt error
  - **Duncan/Xies Galtonian (population/descriptive):** regression summarizes *systematic variability across groups* (following the individualistic paradigm); coefficients describe population differences, not causal mechanisms; error term reflects real within-group variation.
- Absolute size of R^2 should not be overemphasized: low explained variance is expected given substantial within-group variability; While, under the holistic paradigm, as already remarked, the expectation would be that far more of pop- ulation variance than this should be capable of being systematically accounted for, under the individualistic paradigm what is perhaps most remarkable is that regression analyses are usually able to show up some systematic effects despite the fact that the data being analysed being derived by a hopeless jumble of human actors all engaging to some degree in idiosyncratic behaviour as a function of numberless distinctive features of their histories and personalities. error expresses heteroegeneity

8. The limits of statistics: causal explanation

Question 30

Regression and subject-matter knowledge

- For regression to be correctly specified, hence to serve for causal purposes, **subject-matter input** is required:
  - **Which variables to include** (relevant causes, confounders). This rules out **OVB**.
  - **Causal ordering** (which are causes, which are effects).
  - **Functional form** (linear, quadratic, interactions).
  - **Error structure** (independence, homoscedasticity, autocorrelation).

Illustration: Hookes Law

- Regression quantifies this physical law.
- High R^2 expected; error = measurement noise.
- Coefficient has a **life of its own**, representing a specific property of springs.
- This reflects the **Gaussian conception of regression**: law-like causal relation, theory built into model.

Question 31

Limits in sociology

- In sociology, such **Gaussian application is rarely possible**: no physical laws comparable to Hookes Law.
- Issues of **external validity, replicability and parameter stability**. Regression coefficiens may not replicate across **places and times**.
  - Ex parameter change: **If education distribution changes** (if, say, educational differences among ethnic groups were to be narrowed), **coeff of educ on earnings falls**.

Question 32

From regression to social processes

- For **Freedman, understanding the processes generating the data is necessary for proper model specification**.
- For many sociologists, however, **revealing such processes is the very aim of causal explanation** in sociology **independent of statistics.**
- **This addresses the critiques to variable sociology**:
  - **Explaining mechanism, that is, social processes underlying statistical associations in terms of individual action is required to prove causation**.

Question 33

Potential outcomes framework

- Growing
- **Idea**: Under this alternative conception, **causation**: ch**ange that is produced in a dependent or outcome variable of interest as the result of an intervention (treatment)**. A causal effect is **the average difference found in the outcome variable as between randomly selected experimental units that receive a treatment and those that do not** (ATE).
- **Assumptions**
  - **Assignment** is **random**, avoids self selection, ensures balance
  - **Treatment** is **manipulated**
  - **Requires the counterfactual claim**: T group = C group if not treatment.
  - Example rct: same cv change gender, see if bias exists
- **Advantages**
  - **High internal validity**
  - **Simple** (average comparison)
  - **Allows to manipulate reality** in order **to explore**.
  - good RCT does not need control variables

Question 34

Extensions and criticisms

- Many think potential outcomes approach must be taken as a way to concice the ideal Gold standard.
- Experiments have Three difficulties in sociology:
  1. Focuses on **effects of causes** (hume, inductive), not **causes of effects** (Aristotle, deductive). This implies a very **different orientation** from starting out from effects, as, say, established population regularities, and then seeking a causal explanation of them. The interventionist approach is not **often applicable**, because many big questions are those about the causes of effects (eg what leads fertility to decline).
  2. **External validity**, same criticism as regression, **for results to be externally valid the researcher needs to show processes generating the effect to prove causation, and the conditions under which such mechanisms operate**.
  3. Fundamental issue: mechanisms must be expressed in terms of individual action. However, **this requirement then comes into direct conflict with Hollands maxim, basic to the potential outcomes approach, of no causation without manipulation.**
     - Statements:
       (a) She did well because she was coached. → valid (intervention).
       (b) She did well because she is a woman. → a bit problematic (non-manipulable).
       (c) She did well because she studied. → problematic: voluntary action, informed choice.
     - Highlights tension: **voluntary action and non-manupilable attributes** do not fit easily with no causation without manipulation.

4. Other issues:
   - **Outcomes often not prespecified**
   - **Pre-registration scarce**
   - **Sample size determination must be transparent, with power calculation**
   - **Random allocation sequence almost never documented**: over-reliance on convenience (students) vs representativeness.
   - **Attrition and compliance often missing.**
   - **External validity & scalability discussed only superficially**. Internal validity is prerequisite for external validity.
     * Participants = Does effect hold across groups (e.g., students vs. general public)?
     * Settings = Does effect hold in different contexts (lab vs. field, US vs. Europe)?
     * Treatments = Does effect hold for similar but not identical manipulations?
     * Outcome measures = Does effect hold for different operationalizations of the same concept?
   - **Replicability**: concerns about null findings, selective reporting, reproducibility.
   - Social desirability bias
   - null findings, open source data are important

9. Causal explanation through social mechanisms

Question 35

From laws to mechanisms

- Earlier view: causal explanation = showing phenomena followed from general **covering laws of deterministic kind**.
- Modern sociology: To make population regularities TRANSPARENT, two requirements:
  1. They must be 1) **explanatory** (**causally adequate to generate the regularities**), in the sense that they can be explained to emerge from individual informed choices and interaction.
  2. Their operation must be **open to empirical test**. Not about what mechanisms **could** produce the effects observed, more about whether proposed mechanisms *actually* operate in specific cases.
- Mechanism explanations **do not equal technical advance** (e.g., adding intervening variables or diagrams). Input is sociological, output is **generalised narratives of action and interaction** underlying regularities.
- This addresses the main critique to variable sociology.

Question 36

Two approaches in the development of mechanism-based explanations

- **First approach: Analytical sociology**
  - Creates a **toolbox of mechanisms** that operate in social life.
  - Advantage: **systematic development** through the same or similar mechanisms being found to operate across a range of different substantive domains, eg *Matthew-effect*.
  - Danger: **Focus may shift from explaining population regularities to illustrating mechanisms (that is, more focus to effects of causes)**; or, one could say, attention centres simply on causal adequacy, in the sense previously indicated. If one focuses only on cases specially selected so as to best illustrate their application, empriical test requirement is lost!
- **Second approach: Population science congruent**
  - **Starts with probabilistic regularities** that remain opaque (cause of effects problem).
  - Aim: **make visible** how they derive from action and interaction under given conditions.
  - **Risk: mechanisms may appear ad hoc**, **but** multiple hypotheses **can be tested** empirically!

Question 37

Testing mechanisms

- Three strategies:
  1. **Direct observation**: **case studies, interviews** with youth/-parents to observe continuous processes.
  2. **Indirect observation**: **look for implied regularities** (e.g. RRA predicts kinked parental influence),**hypothetico-deductive**, Popper. **Parental background should weaken once children have reached an educational level that likely avoids downward mobility**.
  3. **Experimental/quasi-experimental**: interventions (e.g. Italy RCT on information/advice to students). Under the RRA theory, while some reduction in secondary effects is expected, such effects will largely remain (info is not the main channel).
- Important: mechanisms should be tested **in as many ways as possible**, results compared across strategies. and greatest weight has then to be given to how far results from different tests do or do not fit together (**crossword-puzzle model**).

Question 38

Issue of mechanism-based explanation

- **Objection: infinite regress.** Seeking generative mechanisms underlying regularities risks endless regress: each mechanism can itself be explained by a more primitive mechanism.
- And If mechanisms appeal to **social norms**, questions remain: why these norms, why followed or challenged? Until questions of this kind are answered, it could be held that black boxes clearly do exist
- If mechanisms appeal to **rational action**, a stopping point can be reached: rational action can serve as its own explanation, as rational action is its own explanation. This would prove the **hermeneutic** requirement.

10. Economics, Sociology, and Public Role

Question 39

Sociology vs. Economics, Demography, and Epidemiology

You should now know how sociology works (pop science)

Sociology vs. Economics

- similarities: individualistic paadigm, rational action
- **Economics:** Since the **Pareto turn**, established as a **separate science**, based on **deductively derived rational choice axioms claiming objective correctness**. Empirical findings illustrate or quantify theory, rarely test it. No **falsificationist approach**.
- **Behavioural economics and new economic thinking:** Attempts to bridge the gap with sociology; e.g., studies on *income and wealth distribution* (Atkinson, Piketty).

Sociology vs. Demography

- **Demography:** Already a **population science**, with methods and issues close to sociology (e.g., residential choice, migration).

Sociology vs. Epidemiology

- **Epidemiology:** Also a **population science**, sharing methodological similarities with demography and sociology, especially in studying **population-level regularities** (e.g., disease distribution, health inequalities).

Sociology vs. History

- Historical studies are *ex post* (singular events), but acceptable when searching for **population regularities** in past societies.

Question 40

Weber on Sciences Public Role

- **Objective:** Provide **clarity** about what
  - Science can: **facts**, **analysis**, **theory**.
  - Science cannot
- **ScienceValues link:** Clarifies:
  - To what extent **values** can be realized under given **conditions**.
  - By what **policies**.
  - With what **unintended consequences**.
- **Misuse:**
  - Politicians: claim **facts are on our side** (via **selectivity** or **distortion**) → inherent to politics.
  - Greater danger: **scientists exploiting authority** to present **political values** as **scientific truth**.
- **Vocation of science:** Responsibility to:
  - Maintain **clarity**.
  - Accept **inconvenient findings**.
  - Resist **ideological misuse**.
  - Treat science as a **vocation**.

According to Weber, science must remain **objective**, with personal value judgments excluded.

- The **prophet** and the **demagogue** have no place on the academic platform.
- Science should be **free from presuppositions**.
- The teachers primary task is to train students to recognize **inconvenient facts**, i.e., facts that contradict their political or party convictions.

Historically, this stance represented a **sharp opposition to Marx**, whose scientific approach was explicitly grounded in political commitments.
In Weberian terms (which I personally share to be fully honest and transparent), ńreflexiveż and subjective knowledge cannot be scientific by definition. From a Weberian view, reflexive/public sociology is not true science because it is value-driven and subjective.

The modern Weberian view holds that sociologists, like everyone else, inevitably have their own values, but they must keep these separate from their scientific work. One does **not need to know in advance which is the truth**.
By contrast, a Marxist position assumes that the truth is already known and that research serves to demonstrate it. This often leads to the **selective use of evidence**, whether consciously or unconsciously.
Two examples from my personal experience illustrate this:

- Research on work, unions, and industrial relations.
- Research on migration.

In both cases, researchers were convinced of the positive effects and tended to discard negative evidence.

11. Appendix: Agent-Based Modelling (ABM)

Definition (Central Idea)

- **Structure:** a computational (or conceptual) model where individual agents follow simple rules, interact locally, and through these interactions produce emergent macro-level patterns. IN VITRO

Key Assumptions of ABM

1. **Autonomy:** The system is **not modeled as a globally integrated entity**; **macro-patterns emerge bottom-up**.
2. **Interdependence:** Agents **influence** each other **directly** or via the **environment** (e.g., neighborhood).
3. **Properties and rules:** Agents have **attributes** and follow **constraints**. Models aim to 'explore the simplest set of behavioral assumptions requied to generate the macro pattern of interest'
4. **Adaptation and backward-looking:** Agents **learn from past behavior** (e.g., Bayesian updating).

Approaches to ABM

- **KISS (Keep It Simple, Stupid!): Low-dimensional realism**, abstract, simple models.
- **Keep it as simple as suitable: high-empirical realism** models represent phenomena in **many dimensions**. Detailed population and context representation.

Side note: Trade-off between **parsimony vs realism**. Economics → **parsimony**; Sociology → **realism**.

How to build and ABM

- **Empirical sources** for an ABM:
  - **Population characteristics:** Census (e.g., U.S. IPUMS), administrative records, surveys.
  - **Behavioral data:** Revealed preferences (observed choices), stated preferences (survey experiments).
  - **Contextual data:** Geography, networks, organizations, institutions.
- **How do we make them take decisions?**
  - Estimate probability of action conditional on covariates (e.g. "move house this year= yes/no").
  - Take action with highest expected utility, better if embed values believs ad norms
  - Heuristic rules
- Generate **stochastic variability** with multiple runs
- **Calibrate the model** through Monte Carlo sampling over parameters.
- **Test** multiple specifications

Advantages

- **Bottom-up perspective for micro-macro problems:** System described from constituent units. "Grow it to explain it!". You can see *in vitro* how large regularities are grown.
  - Example: Insightfull to provide mechanism based explanations grounded on feedback loops, contagion, tipping points (small changes trigger large shifts) etc
- Allows to identify hidden assumptions and mechanisms
- **Micro-founded formalization:** ABM uses **computational simulations**, not only **closed-form solutions** to explore dynamics out of the reach of pure mathematical methods.
- Internal validity controlled by researchers
- I believe some risk of overfitting