Groupe de recherche en Droit, Economie et Gestion
Gredeg

UNIVERSITÉ CÔTE D'AZUR

THE TRUST GAME: A HISTORICAL AND METHODOLOGICAL ANALYSIS AT THE FRONTIER OF EXPERIMENTAL AND BEHAVIORAL ECONOMICS

Documents de travail GREDEG
GREDEG Working Papers Series

Nicolas Camilotto

GREDEG WP No. 2026-11
https://ideas.repec.org/s/gre/wpaper.html

The Trust Game: A Historical and Methodological Analysis at the Frontier of Experimental and Behavioral Economics

Nicolas Camilotto[1]

GREDEG Working Paper No. 2026-11

Abstract

This paper provides a life-cycle analysis of the Trust Game, using its trajectory as a lens to clarify the boundaries between experimental and behavioral economics. We first trace its 1995 creation by Berg et al. as a challenge to calculative trust paradigms. A bibliometric study then maps its diffusion, revealing two divergent paths in economics: one, rooted in experimental economics, prioritizes measurement; the other, in behavioral economics, theory-testing. These paths differ in methods and validity standards, constituting an epistemic divide that illuminates the fields' evolving relationship.

Keywords

Trust; Trust Game; Experimental Economics; Behavioral Economics

JEL Classification

B2; B4; C9

1. Introduction

The concept of trust, which has attracted growing interest in the social sciences since the 1980s, has been largely shaped within economics through experimental methods (Camilotto 2023; Tazdaït 2008). At the core of this approach is the *Investment Game*, developed by Joyce Berg, John Dickhaut, and Kevin McCabe in 1995. It became widely known as the *Trust Game*, as it came to represent the "canonical protocol and structure" (Smith 2020, 14) for observing and measuring trust in experimental settings. This game is regarded as one of the "games that inaugurated entirely new directions in experimental research on economic behavior and human sociability" (Smith 2020, 7). It has since come to form, alongside the *Dictator Game* (Forsythe et al. 1994) and the *Ultimatum Game* (Güth, Schmittberger, and Schwarze 1982) a triptych of experimental tools extensively used to investigate social preferences (Camerer 2003).[2]

Despite a considerable body of recent literature on the Trust Game, including meta-analyses (Balliet and Van Lange 2013; Johnson and Mislin 2011), studies on its limitations (Alós-Ferrer and Farolfi 2019) or distinctiveness (Smith 2020), no study has yet examined the methodological dimensions of this experiment through a historical lens. Such a perspective brings into relief two

[1] Université Côte D'Azur, CNRS, GREDEG, France.

[2] The *Dictator, Ultimatum*, and *Trust Games* are often used in combination to provide a comprehensive profile of participants' social preferences in experimental settings (see, for example, Chai, Dorj, and Sherstyuk 2018; Hergueux and Jacquemet 2015; Rustichini and Villeval 2014).

dominant approaches within the economic literature that employs experimental methods. By situating these approaches within a broader historical and methodological perspective, this paper leverages the Trust Game's status to clarify the blurry boundaries between experimental and behavioral economics.

In this paper, we will approach behavioral and experimental economics as two distinct entities. Both originally refer to *scientific revolutions* that took place in economics between the 1970s and 1980s. Experimental economics can be considered a *methodological revolution* that fundamentally challenged the very identity of economics as a non-experimental science (Guala 2010). Its contribution is not limited to the introduction of laboratory experiments; it lies primarily in a radical redefinition of the relationship between theory and data (Svorenčík 2015). Behavioral economics, on the other hand, can be seen as a *theoretical revolution* involving the incorporation of psychological concepts and methods into economics, initially with the aim "to study various empirical deviations of individual behaviors from the predictions of standard models" (Jullien 2018, 119).

As these two movements succeeded, they evolved from *revolutionary* episodes into established fields of research. Behavioral economics emerged as a subfield aiming to move beyond the rational *homo œconomicus* by developing more psychologically realistic models, for which experimentation serves as a key, though not exclusive, source of evidence (Guala 2010). Conversely, experimental economics is best defined as a subfield distinguished by its use of experimental methods, without primarily aiming to model underlying psychological mechanisms. While this distinction appears clear in theory, the practical boundary between behavioral and experimental economics is notoriously blurry.

This ambiguity stems from several intertwined factors, the first of which is historical. Over time, the labels *experimental economics* and *behavioral economics* have been applied to a wide array of research programs. On the other hand, the *experimental economics* label has proven equally expansive, encompassing the early *market experiments* of Edward Chamberlin and Vernon Smith (Guala 2010; Roth 1993) or the German school of bounded rationality around Reinhard Selten (Selten 2003). The case of Selten is particularly illustrative of this taxonomic overlap: while a pioneer of the experimental method, his work is also frequently institutionalized under the banner of behavioral economics, as seen in the common designation of the *Selten School of Behavioral Economic*s (Ockenfels 2010). This dual identity underscores the difficulty of maintaining a rigid distinction between the two fields. Ultimately, in a particularly loose application, the labels behavioral economics and experimental economics are even used interchangeably, further exemplifying this confusion.

A second source of ambiguity lies in the significant overlap between the two fields. Experimental economics has evolved from a marginal subdiscipline into a mainstream method employed across nearly all economic fields, notably within behavioral economics itself. A parallel expansion occurred in behavioral economics. As it evolved from a unified American program into a multipolar, international field with independent subspecialties (Truc 2022a, 393), its very success led Erik Angner to remark that "we're all behavioral economists now" (Angner 2019, 1997). This general blurring of disciplinary lines has fostered an extensive common ground marked by constant circulation between the two categories. This fluidity is driven by the fact that many behavioral economists are also active experimentalists, while experimental research frequently incorporates psychological frameworks. As a result, researchers have long struggled to classify themselves into a single field, with many identifying with both communities (Loewenstein 1999, 25).

Faced with the challenge of narrating the history of diffuse programs that are deeply interconnected and have profoundly penetrated economic analysis, the historiography of behavioral and experimental economics has explored two main avenues. The first approach focuses on specific subfields or episodes (Grüne-Yanoff 2016; Svorenčík and Maas 2016); the second is a quantitative one using bibliometric tools to map large fields over extended periods (Truc 2022b). This paper aims to integrate these two avenues into a single framework: which we term a *scientific quantitative microhistory*.[3] This paper centers on a specific experimental game, employing it as a lens through which the broader dynamics of experimental and behavioral economics are explored.

To conduct this quantitative microhistory, our analysis unfolds in two complementary stages, which follow the life cycle of the Trust Game itself. The first part adopts an analytical focus, providing a historical overview of the game's conception (Section 2). It situates its intellectual context and initial objectives while emphasizing its conceptual construction. As the protocol stabilized, however, the historical narrative shifts from the development of a single design to its diffusion. The second and central part of the paper therefore shifts to a contextual focus, employing the tools of quantitative bibliometrics to map this diffusion (Section 3). This allows us to situate the Trust Game within the broader scientific and social dynamics that have shaped its appropriation across different scientific traditions. The analysis reveals an *epistemic divide* within the economic literature on the Trust Game, pointing to two distinct research trajectories.[4] The first, rooted in an experimental approach, examines the social mechanisms of trusting behavior and employs the game as a policy-oriented measurement tool. The second, drawing on behavioral traditions, investigates the psychological foundations of trust with the aim of developing a theory of trusting behavior. Drawing on the philosophy of economics and philosophy of science, this divide will be explored through differences in scientific approach, experimental design, and standards of validity.

2. Trust in the making - From Theoretical Skepticism to Experimental Validation

Although trust had already been an integral part of psychology and sociology for decades, the concept of trust first entered the field of economics in the 1980s (Camilotto 2021). To understand the intellectual environment from which the Trust Game emerged, this section traces its conceptual origins. Our first step will be to analyze the early efforts to incorporate trust into economics, most of which sought to reduce it to self-interest under the framework of *calculative trust* (2.1). We will then show how experimental approaches, drawing on insights from game theory, directly challenged this paradigm. This process culminated in the Trust Game (2.2), which was developed by Berg, Dickhaut, and McCabe in 1995. The robustness of the game's results (2.3) solidified its status, transforming it from a scientific curiosity into a new experimental paradigm that would shape the economic study of trust for decades to come.

[3] Microhistory is a historiographical approach that concentrates on small, well-defined units of research in order to challenge, test, or add nuance to large-scale historical narratives (See for example Levi 1992; Revel 2010; Magnússon and Szijártó 2013).

[4] Here, the term *epistemic* is used to denote differences in research methodology and epistemological assumptions within a field and should not be conflated with Foucault's notion of *episteme*, which is much broader and historically situated (Foucault 1989).

2.1 The Paradigm of Calculative Trust

Economists did not unanimously embrace the introduction of the concept of trust into economics at the end of the 1980s, being described as "intellectually intrigued but not scientifically convinced"(Laurent 2012, 25).[5] Indeed, from the late 1980s until the late 1990s, a considerable body of economics literature sought to conceptualize trust by simplifying it to a matter of interests. This specific form of trust, also known as *calculative trust* or *encapsulated trust* (Laurent 2012, 28; Tazdaït 2008, 33), originated from the theory of *incomplete contracts*. This literature may be framed in terms of three distinct approaches: *property rights*, *transaction costs*, and *self-enforcing implicit contracts* (Chaserant 2008; Lorenz 1999).

Oliver Williamson's work on transaction costs perfectly reflects the initial skepticism that surrounded the concept of trust as it slowly entered economic analysis. He argued that, in many situations, the notion of trust should be replaced by that of risk — a concept familiar to economists and central to what Williamson called *calculativeness*. According to Williamson, *calculativeness* characterizes the reasoning of economic agents and should occupy a central place in economic theorizing, particularly as a way to extend the economic approach to topics traditionally explored by other social sciences (Williamson 1993, 453). In this view, given that *calculativeness* covers such a broad range of human exchanges, trust—if it occurs—is largely reserved for "very special relations between family, friends, and lovers"(Williamson 1993, 484), while in economic interactions it is deemed unnecessary and may even "[muddy] the (clear) waters of calculativeness" (Williamson 1993, 471). This emphasis on calculative reasoning is emblematic of a broader intellectual movement that sought to formalize the governance of transactions without invoking trust explicitly.

This project took several influential forms. The property rights approach, for instance, centered on interorganizational trust and stressed that trust is essential only when property rights are distributed sub-optimally (Grossman and Hart 1986). Consequently, the focus shifted from trust to a normative debate about how property rights should be distributed to ensure efficient production (i.e., without trust in interorganizational relations). One other major contribution to this field comes from economist and game theorist David Kreps, who sought to incorporate the notion of reputation into economic theory.[6] It was in this context that he created the first game to be called the Trust Game, though it differs from the experimental Trust Game as it is commonly understood today.

In this game, illustrated in Figure 1 and described by Kreps as a "one-sided version of the prisoner's dilemma game" (Kreps 1990, 101), player A must first decide whether or not to trust another player B. If A decides not to trust, both A and B receive nothing. If A decides to trust B, then B has two options to choose from. B can either honor the trust invested by A, in which case both players receive $10, or choose to betray A's trust, receiving $15 while A loses $5. Kreps uses this conceptual framework to address what he terms the "missing pieces" of rational choice theory (Kreps 1990, 131), highlighting, for example, the role of reputation, which he argues could help overcome the inefficient outcomes to which rational choice theory appears confined. Kreps' game, described by Alós-Ferrer and Farolfi (2019, 3) as the "original Trust Game", exemplifies the essential characteristics that are almost always present in experimental investigations of trust in economic interactions:

[5] Translations from French to English are mine.

[6] The reputational approach, and in particular the work of (Kreps 1990), can be seen as an integral part of incomplete contract theory (Lorenz 1996, 491).

1. The act of granting trust should be voluntary.
2. There is a delay between the decision of the trustor (the one granting trust) and the trustee (the one receiving it).
3. The trust decision of the trustee only occurs if and when the trustor decides to trust.
4. The trustor is put in a worse position when their trust is not respected than when they choose not to participate in the situation.[7]

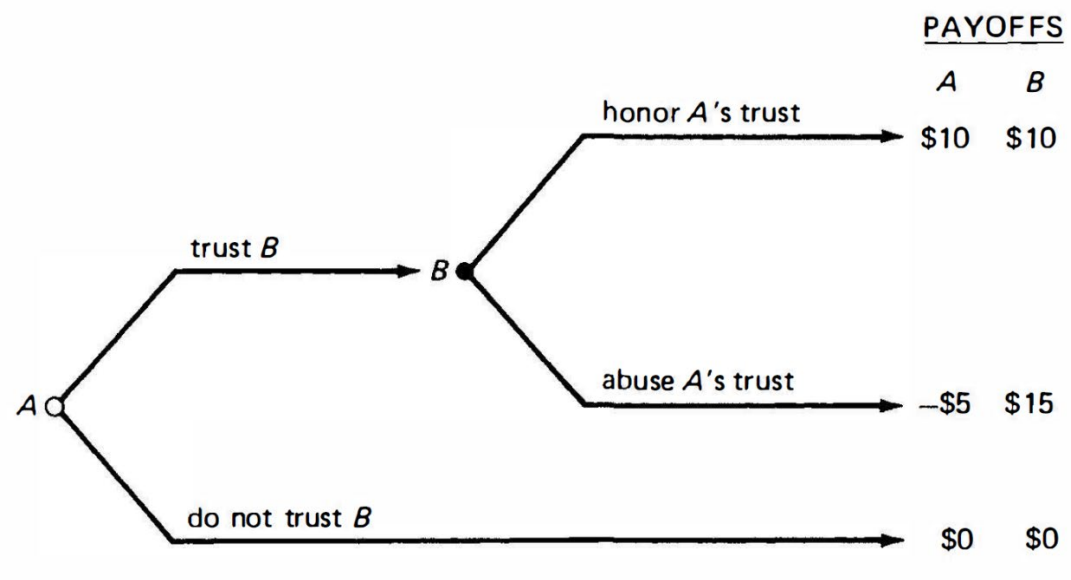


Figure 1 Kreps Trust Game (Kreps 1990, 100)

The development of *calculative trust* theory ended abruptly, as experimental studies using the Trust Game demonstrated that trust cannot be fully encapsulated within self-interested calculations, undermining the theory's premises.[8] Experimental approaches have enabled the discovery and examination of many anomalies, that is, behaviors that cannot be explained by the assumption that agents possess stable, well-defined preferences and consistently make rational choices based on them (Camerer and Thaler 1995, 209). The *Investment Game* is an experimental protocol created in 1995 by Berg, Dickhaut, and McCabe. This protocol is one of many games in which behavioral anomalies with respect to rational choice theory and game theory are observed, challenging the concept of *calculative trust* and transforming the way trust is addressed in economics.

2.2. The Experimental Turn: Inventing the Trust Game

The Trust Game, illustrated in Figure 2, unfolds in the following way: participants are first gathered in a room and then divided into two groups. Half of the group plays the role of trustors and is directed to room A. The other half of the group plays the role of trustees in room B. All participants receive a show-up fee of $10 in one-dollar bills. While participants in Room B must pocket their $10, those in Room A can choose to send an amount between $0 and $10 M_a to a randomly paired participant in Room B. The experimenter then triples the chosen amount $3M_a$ and gives it to B.

[7] These are characteristics that can be initially traced back to Coleman (1988).

[8] The idea that experimental economics has a function to test theories is a recurring theme among experimentalists, philosophers, and historian of economics thought (Guala 2005; Smith 2020; Sugden 2005).

Both A and B know this. Subsequently, the participants in Room B have the option to return a portion of the received amount to their paired counterpart ($(k_b(3M_a))$). The choices of participants in Room A and B remain anonymous through the employment of a double-blind procedure. This protocol guarantees that no participant observes the decisions of another, while also preventing the experimenters from observing any individual's choices.

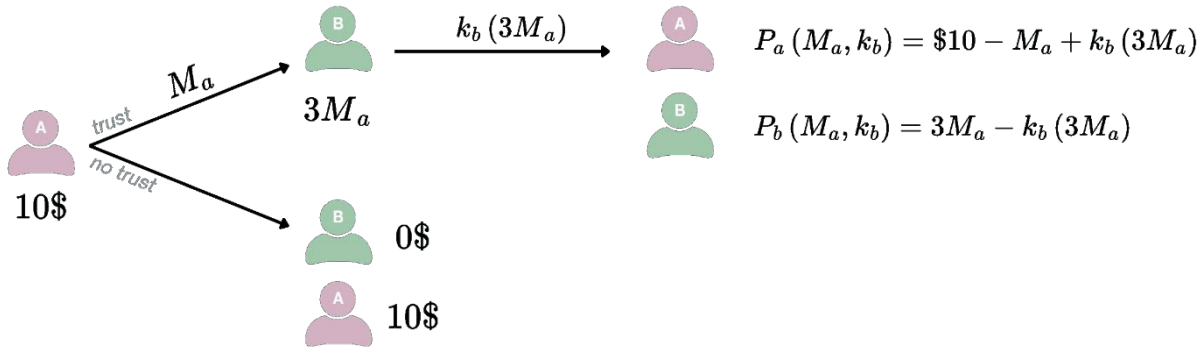


Figure 2 Berg, Dickhaut, and McCabe (1995) Trust Game

The Trust Game draws inspiration from the previous trust-based games; however, this experiment introduces innovative characteristics that explain the popularity of the game in comparison to its predecessors. Firstly, the amount of money sent is tripled, indicating that it is not a *zero-sum game* where the loss of some equals the gain of others; rather, it produces a higher social gain through cooperative behavior. Furthermore, the Trust Game is able to measure varying levels of trust, unlike the previous trust-based games, which forced binary choices. A person has the choice to make a greater or lesser investment in their partner's trustworthiness, while their partner can also choose to honor or betray the trust placed in them to differing extents. Within this game, the amount sent by participants in Room A is intended to measure trust, while the amount returned by participants in Room B is intended to measure trustworthiness. This protocol therefore establishes quantitative levels of trust and trustworthiness, making it easy to compare different sessions of the experiment. The paper by Berg, Dickhaut, and McCabe is a prominent illustration of this ability to make comparisons.

Berg and his co-authors sought to challenge the "fundamental assumption in economics [...] that individuals act in their own self-interest" (Berg, Dickhaut, and McCabe 1995, 122). To accomplish this, the authors designed an experimental protocol that deliberately excluded the very conceptual mechanisms central to the *calculative trust* paradigm, such as "reputation from repeat interactions, contractual precommitments, and potential punishment threats" (Berg, Dickhaut, and McCabe 1995, 123). The initial experiment consisted of six sessions, divided into two sets. The first three sessions served as a control condition, designed to investigate trust development in the absence of these mechanisms.

The final three sessions formed their treatment condition, in which the authors sought to measure the effect of what they call "social history" (Berg, Dickhaut, and McCabe 1995, 123) by allowing participants to review the results of the first three sessions before playing. This treatment aimed to test hypotheses based on an evolutionary perspective of trust (Hirshleifer 1977; Selten 1989; Güth, Ockenfels, and Wendel 1993). This perspective argues that trust arises because it maximizes *genetic fitness*, even when myopic self-interest would suggest cheating. It also posits that social history provides common information about trust within an organization and may further reinforce individuals' predispositions to trust (Berg, Dickhaut, and McCabe 1995, 124).

During the first three sessions, only 2 of the 32 subjects in Room A chose not to send anything to their counterparts in Room B. The average amount sent was $5.16. On average, participants in Room B returned $4.66 to their Room A counterparts. Out of the 28 Room B recipients who received more than $1, 11 of them returned a higher amount than their partners in Room A had sent. Although the expected gain for individuals who have chosen to trust is negative, it is nonetheless evident that the observed outcome diverges significantly from what rational choice theory would anticipate—namely, the impossibility of any exchange. According to the authors, this result—achieved without using typical rationality-based mechanisms to enforce trust—serves as evidence that "self-interest alone cannot explain [trust]" (Berg, Dickhaut, and McCabe 1995, 137). In the second set of sessions, participants were provided with the results from the first round. In Room A, the average amount sent increased slightly to $5.36, without this increase being statistically significant. However, in Room B, the increase was statistically significant, with players sending an average of $6.46 back to Room A (an increase of $1.[9] between the two series of sessions).8 For the authors, this treatment posed a direct challenge to the evolutionary approach. This perspective predicts that information about past mixed outcomes should lead to less trust and reciprocation, as rational agents learn to be more cautious. Yet, the experimental results pointed in the opposite direction. The increased reciprocation from Room B was so significant that it created a positive expected gain for those who chose to trust, directly contradicting the prediction that social history would diminish, rather than enhance, cooperation.

These results represented a direct and compelling challenge to the prevailing paradigm of *calculative trust*. Providing powerful, quantifiable evidence that significant levels of trust and trustworthiness could emerge even in sterile laboratory conditions was more than a mere documentation of an anomaly; it was the establishment of trust as a legitimate and observable economic phenomenon.

2.3. From Anomaly to Robust Phenomenon

This initial finding, however, immediately raised a critical question: was the behavior observed by Berg and her co-authors a fragile artifact of their specific design, or was it a genuinely robust feature of human interaction? This section demonstrates how decades of subsequent research have confirmed the latter, establishing the resilience of the Trust Game's core result as one of its most defining characteristics.

Indeed, much of the research following the original experiment can be understood as a series of *stress tests* designed to probe the limits of this phenomenon. By intentionally fostering selfish behavior, the initial protocol already conveyed the message that "it is OK not to send money, and OK to keep any money received." (Smith 2020), a logic that some experimental designs pushed even further.[10]

For instance, Ortmann, Fitzgerald, and Boeing (2000) suggest enhancing the *social history* of the Trust Game by presenting the results of an initial control session and by administering a questionnaire to participants in Group A before the actual experiment.[11] This two-part protocol

[9] The statistical significance of this result was evaluated with a one-sided Wilcoxon rank-sum test (r = 776) and found to be significant at the p = 0.1 level.

[10] These kinds of experiments, which use extreme protocols specifically to test the robustness of established results, are usually referred to as "boundary experiments" (Smith 1982, 942; Kaplan 1964, 150).

[11] The questionnaire asks Room A participants four questions: "How much money do you think you will send?"; "How much money will your Room B counterpart receive if you send this amount?"; "How much

has two objectives: first, to emphasize even more clearly to subjects in Group A that sending money to Group B is not profitable; second, through the questionnaire, to encourage *rapid strategic reasoning*. To the authors' "great surprise", however, the experimental protocol failed to diminish the amount invested during the initial phase (Ortmann, Fitzgerald, and Boeing 2000, 82). McCabe and Smith (2000) conducted another skeptical investigation of the Trust Game. Their experimental protocol involves dichotomizing the participants' choices, where they can only choose between two highly contrasting actions. The choices in this game are illustrated in Figure 3.

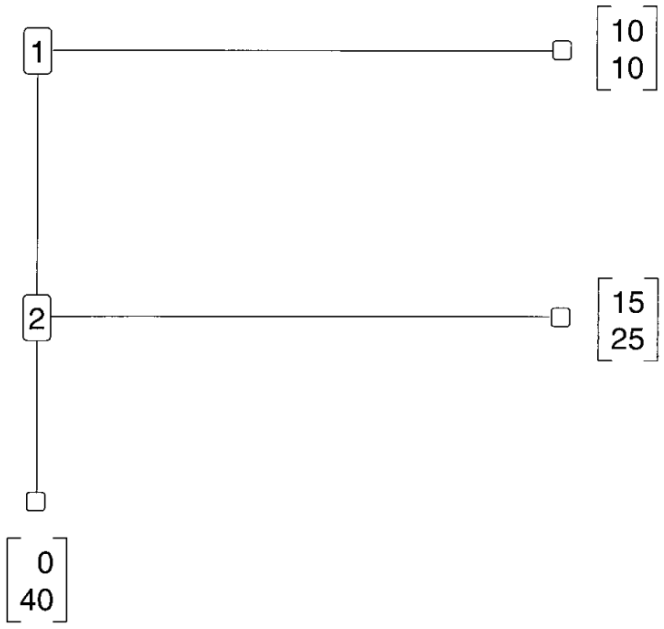


Figure 3 The Trust Game by McCabe and Smith (2000, 3778)

Participants in Group A are presented with two options: either to send nothing, in which case each player keeps their show-up fee of $10, or to send this show-up fee to participants in group B. If group A participants choose to send, the corresponding participants in Group B receive $30 (*i.e.*, triple $10) and then face two options. The first option is a relatively *fair* solution, in which the participant holds on to $15 and returns $15 (to which their $10 show-up fee is added, resulting in a profit of $25). The second option is a selfish solution, in which the participant keeps the entire $30 and their $10 show-up fee, resulting in a profit of $40. The game's outcomes led the experimenters to conclude once again that "remarkably and surprisingly" (Smith 2020, 8) the frequency with which subjects engaged in cooperative behavior was high enough that, on average, both players' payoffs increased relative to the payoffs in the self-interested equilibrium.

Trust in the Trust Game is a sturdy phenomenon, enduring even through experimental protocols "designed strongly to encourage self-interested action" (Smith 2020, 4). This sturdiness was ultimately confirmed on a massive scale by the meta-analysis of Johnson and Mislin (2011). This analysis includes 162 replications of the Trust Game with over 23,000 participants from 35 different countries. Among all the experimental variations studied, only two factors significantly affect the amount sent from Group A: the random payment method (paying subjects based on random selection rather than paying each subject for the amount earned by playing the Trust

money do you think will be returned to you?"; "How much money would you return if you were in Room B?" (Ortmann, Fitzgerald, and Boeing 2000, 83).

Game) has a negative impact on the amount sent, while playing with real players (and not with simulated counterparts) has a significant positive impact on the amount sent. Three variables significantly decrease the amount returned from Group B to Group A: increasing the rate of return (varying the multiplier that determines the amount sent to trustors), playing exclusively with students, and allowing players to assume both roles within the same session (Johnson and Mislin 2011, 873).[12]

Born as a tool to challenge the self-interest paradigm, the Trust Game evolved from an experimental curiosity into a powerful scientific instrument, an evolution driven by the exceptional robustness of its core behavioral finding. While minor variations would emerge, the core methodology of the Trust Game saw no further major developments after its inception and it would progressively establish itself as the canonical protocol for studying trust.[13] As the protocol stabilized, the historical narrative thus changes in nature. It is no longer a story centered on the conceptual development of a single protocol, but one of its diffusion and appropriation across different scientific communities. To trace these divergent uses, and to employ the Trust Game as a lens through which a fundamental divide becomes visible, our historical analytical approach must evolve. While a qualitative examination was ideal for the game's conception, mapping its widespread diffusion requires a shift to large-scale quantitative analysis to navigate this complex landscape of applications.

3. A Common Language, Different Dialects: The Divergent Applications of the Trust Game

This section builds on a quantitative perspective to examine the Trust Game literature. We first present our corpus and methods (3.1) before conducting a network analysis (3.2) that will shed light on the different uses of the game by various scientific traditions for distinct purposes. We then focus on the traditions rooted in economics, examining them from a methodological perspective and considering three aspects (scientific approaches, types of experimentation, and criteria of validity) that define what we term an *epistemic divide* (3.3).

3.1. Corpus and Methods[14]

The first step of our approach was to identify a relevant corpus to study. To do so, we adopted a keyword-based selection strategy, extracting papers from the *Web of Science Core Collection Database* where the expression "Trust Game" or "Investment Game" appeared in the title, abstract, or keywords. This procedure yielded a total of 1,454 papers published between 1995 and

[12] The other studied variables did not significantly affect the Trust Game outcomes, including the amount at stake, receiver endowments (in some protocols, only trustees receive money at the start of the experiment), strategy method (instead of letting trustors decide how much to send to the trustee during the experiment, they apply a predefined strategy for each scenario), anonymity, and a *double-blind* procedure.

[13] The game's canonical status has not, however, shielded it from methodological critiques. Prominent critiques, such as that of (Ermisch and Gambetta 2006), argue that the *standard protocol* blurs the measurement of trust by introducing confounding effects. While this has led to modified protocols, these variations are typically presented as corrections or refinements of the original (Berg, Dickhaut, and McCabe 1995) design, which remains the essential point of departure for subsequent research.

[14] This paper was made possible by a community that builds and provides open-source tools. Data processing was carried out using Positron and would not have been possible without R packages such as networkflow (Goutsmedt and Truc 2022), Bibliometrix (Aria and Cuccurullo 2017), and Biblionetwork (Goutsmedt, Truc, and Claveau 2021). Additionally, generative AI tools were used to assist with coding and writing, including Gemini 2.5, Claude 4, and GPT-4.

2024. However, it presents two important limitations. First, our findings are necessarily contingent on the breadth and quality of our chosen database.[15] Second, by limiting our search to titles and abstracts, we exclude papers that use the Trust Game without explicitly mentioning it there.

The paper by Karlan (2005) paper serves as an illustrative example of the vulnerability of our approach. Due to the lack of an abstract for this paper in the Web of Science database and the fact that its title, "Using experimental economics to measure social capital and predict financial decisions" does not explicitly mention the Trust Game, it was not included in our corpus despite its importance in the literature. However, it was essential to impose a precise and restrictive criterion to avoid including the entire literature on *trust* in our corpus. Nevertheless, our methodological approach is designed to mitigate such omissions. Our quantitative analysis focuses not only on the corpus papers themselves, but also on the most frequently mobilized references within their bibliographies. This secondary analysis reveals that Karlan's paper is, in fact, one of the most cited references by the very articles we selected. Therefore, while the initial keyword-based selection has limitations, our network-based approach provides a more exhaustive view by capturing both the explicit literature on the game and the key intellectual influences that shape it, even those not directly identified by our keywords.

To explore this corpus, we perform a network analysis using the bibliographic coupling method, which, as illustrated in Figure 4, links documents based on the number of references they share in their bibliographies (Kessler 1963). The greater the number of shared references between two papers, the stronger the link between them.[16]

Bibliographic coupling allows the detection of clusters, that is, sets of references that show a certain cognitive proximity, without assuming the reasons for this proximity. The set of identified clusters provides a macroscopic image of the general structure of the corpus, highlighting the "number of communities [clusters], the density of their links (within and between communities), and the position of nodes and communities in the core/periphery structure" (Goutsmedt 2021, 564). To better address our research question, namely the diffusion and use of the Trust Game across different scientific communities, we opted for a specific network representation in the form of an alluvial diagram. Alluvial diagrams are particularly useful in network analysis as they allow us to visualize the evolution of clusters over time,

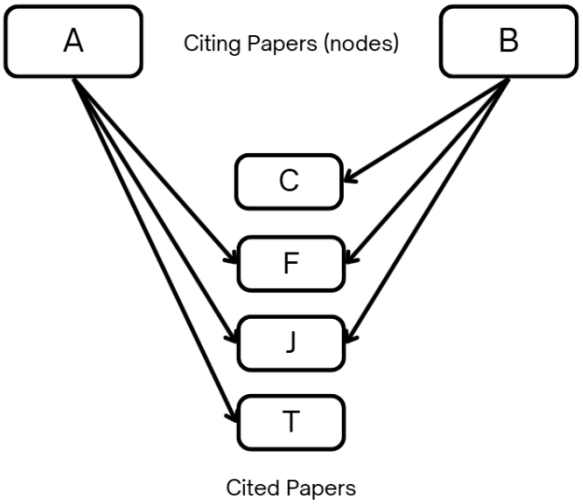


Figure 4 llustration of Bibliographic Coupling: Pallustration of Bibliographic Coupling: Papers A and B are bibliographically coupled because they both cite papers F and J.

[15] The Web of Science Core Collection is a selected catalog of over 21,100 peer-reviewed journals across 250 disciplines. It was chosen for this study as it is a comprehensive and curated database frequently employed in bibliometric analyses within the history of economic thought, providing a level of quality well suited to our research (for example Cherrier and Svorenčík 2018; Edwards, Giraud, and Schinckus 2018; Truc and Jullien 2023).

[16] To normalize and weight these links, the Salton cosine measure is used (Salton and McGill 1983), which divides the number of shared references between two papers by the square root of the product of the sizes of the bibliographies of the two papers. This method has the advantage of taking into account the length of the bibliographies, thus avoiding giving undue prominence to papers with large bibliographies.

highlighting how nodes migrate between communities, how clusters merge or split, and how the structure of the network changes throughout the study period. This approach provides a dynamic perspective on the interconnections and the temporal flow of knowledge between scientific groups (see Appendix A for a detailed description of how the standard network representation is translated into the alluvial diagram).

3.2. Network Analysis of Trust Game Literature

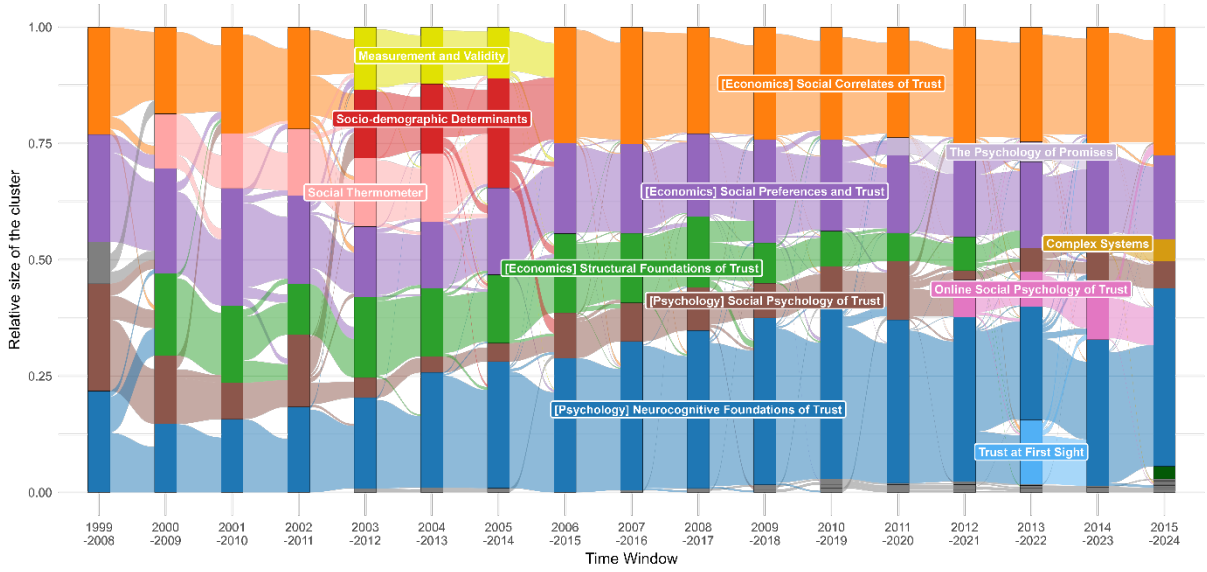


Figure 5 Evolution of the bibliographic coupling network of the Trust Game literature visualized as an alluvial diagram. The visualization shows the changing structure of research clusters across 10-year sliding windows. An interactive version of this visualization is available online: https://019a10aa-000d-6239-b6d1-154738949fe8.share.connect.posit.cloud/.

While assigning names to clusters is inherently subjective, we provide descriptive labels to improve clarity in the following section. Labels were chosen based on a systematic analysis of several factors, including each cluster's thematic focus (key papers and cited references), its disciplinary identity (primary journals and subject tags), and its temporal and geographical distribution (publication dates and author affiliations). To allow readers to assess the relevance of our labeling choices and to explore the underlying data in greater detail, we have developed a complementary online application. This interactive tool allows for a dynamic navigation of the network and provides access to all generated materials. It can be accessed here: https://019a10aa-000d-6239-b6d1-154738949fe8.share.connect.posit.cloud/.[17]

Before turning to the identified clusters, it is instructive to get a sense of the disciplinary landscape of the corpus. Research drawing on the Trust Game is largely concentrated in economics and psychology, while neuroscience, though less prominent, remains part of the landscape. Turning to the temporal dynamics of these disciplinary distributions, we observe that clusters dominated by economics gradually occupy a smaller share of the network, now representing about half of it. Psychology-dominated clusters have correspondingly grown to occupy the other half of the network, largely within the *Neurocognitive Foundations of Trust*

[17] To ensure the reproducibility of the results, the scripts utilized can be found at https://github.com/ncamilotto/trustgame-paper. Furthermore, an archived version of these scripts is available on the Open Science Framework (https://doi.org/10.17605/OSF.IO/ZE2MC) to ensure long-term preservation of the research materials.

cluster. As the alluvial diagram illustrates the proportional distribution of clusters, it necessarily masks the dramatic growth in the absolute volume of publications over time. To illustrate, the corpus grew from 169 articles during the 1999–2008 period to 1,144 during the 2015–2024 period.[18] To examine these cluster more closely, we now turn to an overview of the main objectives and research purposes of the principal clusters.

3.2.1. Neurocognitive Foundations of Trust

The *Neurocognitive Foundations of Trust* cluster brings together research at the intersection of psychology and neuroscience, aiming to elucidate the brain mechanisms and psychological processes that underpin interpersonal trust. A significant part of this cluster focuses on identifying the neural bases of trust and trustworthiness. The foundational study by King-Casas et al. (2005) revealed that a player's reciprocity strongly predicts their partner's future trust. This behavioral dynamic is mirrored by neural responses in the dorsal striatum, a brain region associated with reward and learning. In the same vein, research by Delgado, Frank, and Phelps (2005) demonstrated how perceptions of a partner's moral character modulate the neural reward systems. Prior social and moral information can thus influence trust decisions, potentially reducing reliance on feedback mechanisms and trial-and-error learning in neural circuits.

Another major research axis explores the biological factors that influence trust, notably oxytocin. On one side, the highly cited study by Kosfeld et al. (2005) showed that exogenously administered oxytocin increased trusting behavior. Complementing this, other studies explored the hormonal basis of trustworthiness. Zak, Kurzban, and Matzner (2005), for instance, found that a trustee's endogenous oxytocin levels were associated with their degree of reciprocity. This hormonal axis of research was further broadened by studies on other factors, such as the influence of testosterone on financial risk preferences (Apicella et al. 2008).

In summary, the *Neurocognitive Foundations of Trust* cluster provides a multidimensional view of trust by bridging multiple levels of analysis. It connects the neural mechanics of decision-making in specific brain regions with the modulatory effects of hormones like oxytocin, ultimately showing how these biological factors shape the psychological processes that govern trusting behavior.

3.2.2. Social Psychology of Trust

The *Social Psychology of Trust* cluster is firmly rooted in psychology, investigating the interpersonal, situational, and individual factors that govern the decision to trust and to be trustworthy. The comprehensive meta-analysis by Thielmann, Spadaro, and Balliet (2020) provides a robust framework linking personality traits to prosocial behaviors like trust, demonstrating how these traits are activated by specific situational *affordances* such as vulnerability or the opportunity for reciprocity. Beyond stable dispositions, research explores the heuristics and social cues that guide trust in the absence of concrete information. For instance, DeBruine (2002) demonstrated that facial resemblance enhances trust, pointing to an instinctual mechanism that may favor those perceived as kin or familiar. Similarly, the research by Dunning

[18] While this absolute increase is significant, it must be contextualized within the broader trend of substantial growth in scientific publishing. Hanson et al. (2024), for instance, note that the total number of articles indexed in major databases has grown rapidly in recent years, with the 2022 total being approximately 47% higher than in 2016.

et al. (2014) reveals that trust at *zero acquaintance* is less about an expectation of reward and more a matter of perceived social respect, underscoring the fundamentally social nature of initial trust.

This cluster examines in depth the interactive dynamics of trust. Researchers investigate the calculus of reciprocity, or how individuals assess others' intentions and decide whether to reciprocate trust. Studies like that of Pillutla, Malhotra, and Murnighan (2003) analyze how the attributions we make about others' motives shape our own behavior. A significant portion of this cluster is dedicated to exploring strategies for trust repair. These studies assess the effectiveness of various mechanisms, such as financial compensations (Desmet, De Cremer, and van Dijk 2011), apologies, or voluntary *hostage posting* to signal future cooperative intent (Nakayachi and Watabe 2005). On a broader scale, the cluster examines the social mechanisms that sustain cooperation within groups, such as the power of reputation and gossip in enforcing cooperative norms, as explored, for example by Wu, Balliet, and Van Lange (2016).

Ultimately, the *Social Psychology of Trust* cluster portrays trust not as a static calculation, but as a dynamic and fragile social process. It reveals how trust is built on subtle cues, guided by individual dispositions, shattered by perceived intentions, and carefully rebuilt through social rituals.

3.2.3. Social Preferences and Trust

The *Social Preferences and Trust* cluster is situated within the field of economics, focusing on the rigorous, model-driven examination of non-selfish behavior in strategic interactions. Reflecting a central ambition of behavioral economics, a core theme of this cluster is the development and testing of social preference models. The highly cited theoretical works, such as Fehr and Schmidt (1999) on *inequity aversion*, Bolton and Ockenfels (2000) on *equity* and *reciprocity*, and Charness and Rabin (2002) on *quasi-maximin preferences*, are foundational. These models propose that individuals' utility depends not only on their own payoff but also on the payoffs of others. Within this research program, the Trust Game serves as a controlled environment to generate precise behavioral data that can be used to test and refine formal economic theories of human motivation. The goal is not merely to observe behavior, but to build models that can explain and predict it.

This cluster also examines a class of models that integrate beliefs and intentions, an approach known as psychological game theory (Rabin 1993; Dufwenberg and Kirchsteiger 2004). These theories move beyond outcome-based preferences to formalize motivations driven by expectations, such as the desire not to disappoint a trustor's belief in one's reciprocity. For instance, the concept of *guilt aversion*, which is explicitly tested by papers within the cluster like Ellingsen et al. (2010), posits that a trustee returns money not because they care about fairness *per se*, but because they do not want to disappoint the trustor's expectation of a return. Similarly, models of reciprocity (Isoni and Sugden 2019) explain trustworthiness as a desire to reward a kind intention (the trustor's initial investment).

A critical contribution of this cluster is the careful disentanglement of trust from simple risk taking. The highly cited work by Bohnet and Zeckhauser (2004) and Bohnet et al. (2008) introduces and provides evidence for the concept of *betrayal aversion*. This trend of research demonstrates that individuals are more averse to losing money through social betrayal than to losing the same amount of money through an impersonal, random process. This distinction is crucial, showing

that trust is not just a calculation of probabilities but involves a unique social and emotional component. Other papers, such as Sapienza, Toldra-Simats, and Zingales (2013) or Schechter (2007), further this agenda by developing experimental methods to separately measure an individual's risk preferences and their specific beliefs about the trustworthiness of others, thereby isolating the core components of a trusting decision.

Overall, the *Social Preferences and Trust* cluster reflects an economic approach to understanding prosocial behavior. It is defined by its commitment to building and testing theoretical models that can explain deviations from pure self-interest. It uses the Trust Game as a precise instrument to measure behavior and test whether that behavior is better explained by preferences over outcomes (like fairness), by belief-dependent motivations (like guilt and reciprocity), or by a specific aversion to social betrayal.

3.2.4. Structural Foundations of Trust

The *Structural Foundations of Trust* cluster approaches trust from a different angle than the previous clusters. Rather than focusing on individuals' psychological predispositions or innate social preferences, it examines how the very structures of social interactions can create the incentives necessary for trust to emerge and be sustained. Drawing heavily on game theory (Kandori 1992; Kreps and Wilson 1982; Kreps 1990) and evolutionary biology (Nowak and Sigmund 1998, 2005; Trivers 1971), this cluster addresses the fundamental issue of how trust and cooperation can emerge among actors when the environment is appropriately structured.

A key element of this cluster is the idea that the possibility of future interaction radically changes the dynamics of trust. Papers like the influential survey by Dal Bó and Fréchette (2018) on repeated games, and empirical studies by Engle-Warnick and Slonim (2004, 2006) on the evolution of strategies in the Trust Game, show that when individuals expect to interact again, cooperation becomes a long-term investment. In this repeated-game framework, trust is no longer a simple leap of faith, but a rational strategy to unlock future mutual gains. Extending this principle to interactions between strangers, Camera and Casari (2009) show that cooperation can be maintained even in anonymous settings, provided the probability of future interaction is sufficiently high.

Significantly, the alluvial diagram reveals this cluster merging with *Social Preferences and Trust* cluster over the final three time windows. This convergence signals a crucial intellectual synthesis in behavioral economics thought, underscoring a key insight: while institutional structures are essential for creating a scaffold for trust, the robustness of that scaffold is deeply intertwined with individuals' intrinsic motivations, such as *inequity aversion*, *guilt*, or the desire for *reciprocity*.

In essence, the *Structural Foundations* cluster provides the *how* cooperation is possible, while the *Social Preferences* cluster addresses into the *why* individuals choose to cooperate within those structures. The merger represents the integration of these two perspectives: trust is the product of a complex interplay between external incentives (the structure) and internal motivations (social preferences).

3.2.5. Social Correlates of Trust

The *Social Correlates of Trust* cluster is rooted in experimental economics and centers on a fundamental question: how do trust and trustworthiness, as measured by the Trust Game, vary

systematically with individuals' demographic, social, and cultural characteristics? Rather than seeking a model of trusting behavior, this field of research aims to map and understand the differences in trust between groups. It uses experimental data to shed light on real-world economic and social issues, from discrimination (Slonim and Guillen 2010) to the impact of social capital on economic development (Bouma, Bulte, and van Soest 2008).

A major theme of this cluster is how trust operates between different social groups. The experiment by Fershtman and Gneezy (2001) is a paradigmatic example. By running Trust Games between Ashkenazi and Sephardic Jews in Israel, they demonstrated evidence of ethnic discrimination, showing that stereotypical beliefs about a group can negatively influence trust decisions. This work paved the way for numerous other studies on how trust is affected by group identity. Research in this cluster explores a variety of social categories, including gender (Buchan, Croson, and Solnick 2008), religion (Tan and Vogel 2008), and even superficial cues like physical attractiveness or a simple smile (Wilson and Eckel 2006; Scharlemann et al. 2001). These studies collectively show that trust is far from blind; it is deeply shaped by social identity, stereotypes, and interpersonal signals.

Finally, this cluster seeks to connect laboratory measures of trust to a variety of real-world social and historical factors. Research in this vein typically examines how trusting behavior is affected by large-scale variables, such as historical legacies of violence (Cassar, Grosjean, and Whitt 2013) or differences across cultures (Holm and Danielson 2005). Others apply this approach to specific field contexts, exploring for instance the role of trust in the management of community resources (Bouma, Bulte, and van Soest 2008).

Taken together, the *Social Correlates of Trust* cluster highlights the socially embedded nature of trusting behavior. Rather than treating trust as a universal and context-free disposition, this body of research reveals how patterns of trust and trustworthiness systematically reflect social identities, stereotypes, and contextual cues. By linking micro-level behavioral variations to macro-level outcomes such as economic growth, discrimination, and resource governance, this cluster provides a bridge between individual decision-making and the structural forces that govern collective life.

This detailed exploration of the five clusters reveals the Trust Game's remarkable versatility, serving as a common language spoken in different dialects across neuroscience, psychology, and economics. Each community has adapted the game to its own theoretical questions and empirical goals, creating distinct but interconnected bodies of knowledge. Having mapped this broad landscape, our analysis now sharpens its focus. While the distinctions between disciplines are significant, a more subtle fault line runs within economics itself. From the perspective of the history of economic thought, this intra-disciplinary tension is particularly revealing. The following section is therefore dedicated to this *epistemic divide*, examining its premises and methodological roots within economics, rather than the more documented differences between economics and psychology (Davis 2013; Lisciandra 2018).

3.3. An Epistemic Divide

A brief preliminary note is necessary before proceeding with this section. Our network-based approach does not draw a strict boundary between experimental and behavioral economics, an expected outcome given that the bibliographic coupling method captures a wide range of cognitive proximities, reflecting both the inherent overlap of the two fields and their complex cross-influences. Acknowledging this ambiguity, we chose to label our clusters according to their research purposes rather than established disciplinary categories. Yet the distinction becomes

evident in our clusters: the research in *Social Correlates of Trust* clearly aligns with experimental economics, whereas the work in *Social Preferences and Trust* embodies the behavioral approach. Therefore, while acknowledging that neither field is fully reducible to a single cluster, we will treat these two as ideal types representing the distinct scientific approaches at the heart of our analysis.

While a comprehensive methodological account is beyond our scope, we will focus on three areas where the *epistemic divide* is most salient within the Trust Game literature: the *scientific approach*, the *type of experiment*, and the *standards of validity*. We select these three points of friction because they form a coherent analytical cascade, from abstract scientific purpose to concrete experimental practice. They can be understood as interconnected levels of the same fundamental divergence: a competing vision of the primary purpose of experimentation in economics.

3.3.1. Theory-Driven versus Data-Driven Approaches

The shared reliance on the trust game might suggest a common methodological ground. In reality, it is the very site of a fundamental scientific schism: one group uses the game to build and refine theoretical models, while the other employs it as a device to measure a social phenomenon.

The experimental economics approach, embodied by the *Social Correlates of Trust* cluster, relies on an inductive scientific approach that seeks to derive general insights from empirical regularities. To achieve this, it employs a standardized device for measuring a phenomenon and revealing robust empirical patterns across different social categories. The trust game is thus employed as a diagnostic and predictive tool, closely tied to the principles of *evidence-based policy* (Favereau 2021). The objective is to generate robust scientific evidence that policymakers can use, particularly in the field of development economics, as exemplified by the work of research centers like the *Abdul Latif Jameel Poverty Action Lab* (J-PAL)[19]. The Trust Game is an integral part of the toolkit employed by this network of researchers, who are particularly well represented in this cluster (Ahlerup, Olsson, and Yanagizawa 2009; Ashraf, Bohnet, and Piankov 2006; Banerjee 2016; Banerjee, Galizzi, and Hortala-Vallve 2021; Chetty et al. 2021; Heyes and List 2016; Jack 2009; Schechter 2007).

The body of work we associate with behavioral economics is more *hypothetico-deductive* oriented. Recognizing from the outset that documenting empirical anomalies alone could not displace the neoclassical paradigm, proponents of behavioral economics understood they needed formal models to explain these deviations (Truc 2022a, 416). Experimentation, in this context, is the primary tool used to forge and validate the components of these new theories. It is not an inductive search for patterns, but a targeted method to test specific psychological mechanisms that can serve as the foundations for a more realistic economic theory. While influencing policy is not the primary ambition of this approach, both private and public actors can nonetheless harness the results of behavioral economics. Psychologically informed interventions exploit these mechanisms to encourage desirable behaviors, with nudges representing one well-known example.

[19] The J-PAL is a global research center based at the Massachusetts Institute of Technology (MIT) that aims to reduce poverty by ensuring that policy is informed by scientific evidence. It conducts randomized controlled trials and other evaluations to assess the effectiveness of social and economic programs worldwide.

In essence, the two approaches use the trust game to answer two fundamentally different questions: *Who is trustworthy?* versus *Why are people trustworthy?*. The first question leads to a science of measurement and prediction, oriented toward policy intervention. The second fosters a science of causal explanation and mechanism design, aimed at theoretical refinement. This fundamental difference in the research question inevitably shapes the research method, manifesting in the design of the experiments themselves.

3.3.2. Experimental Settings as Scientific Commitments

The landscape of economic experiments is defined by a fundamental tension between the precision of experimental control and realism. Following the taxonomy developed by Harrison and List (2004), at one end of this spectrum are *conventional laboratory experiments*, conducted in a controlled university setting, typically with student participants. Their strength lies in the high degree of control, which allows researchers to precisely manipulate variables and isolate clean causal relationships. At the opposite end are *natural field experiments*, which take place in real-world environments where participants are unaware that their behavior is being studied. This method provides unparalleled realism, but the loss of control makes it more difficult to disentangle cause and effect. To navigate this trade-off, economists have developed innovative hybrid methods. An *artefactual field experiment* is similar to a *conventional laboratory experiment* but uses a nonstandard subject pool. A *framed field experiment* builds on an *artefactual field experiment* by incorporating elements of the field context, such as the commodity, task, or information available to participants.[20]

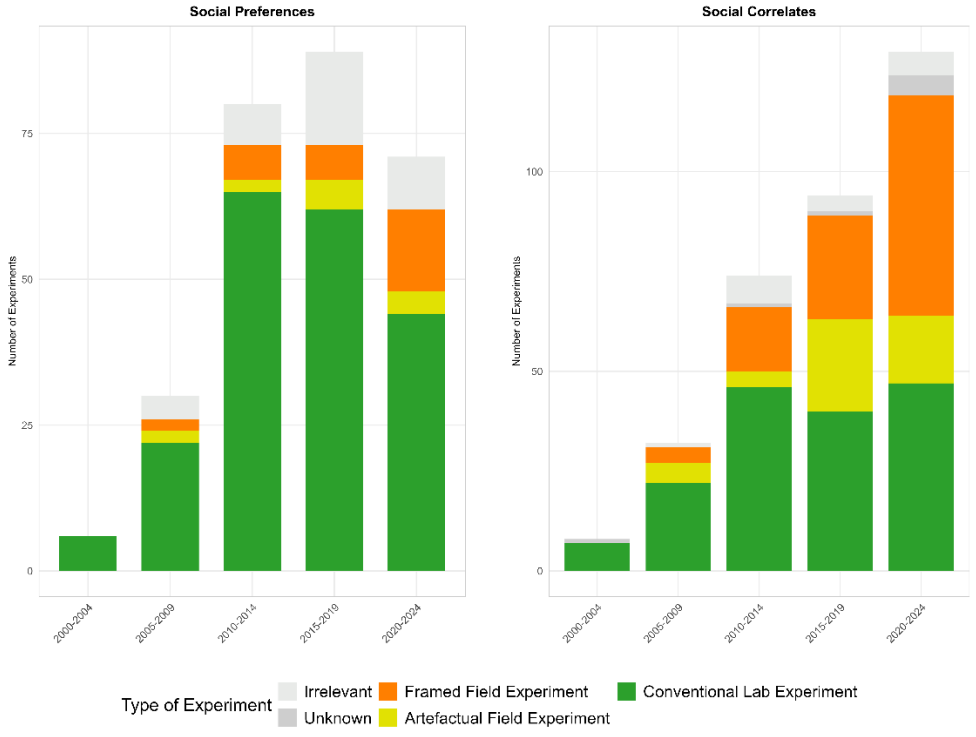


Figure 6 Evolution of the number of experiments by type over 5-year windows for two clusters: Social Preferences and Trust and Social Correlates of Trust.

[20] It is worth noting that this taxonomy, like all such classifications, is an imperfect guide. Some studies defy easy categorization, while a single label may encompass a wide variety of distinct experimental practices.

Figure 6 provides striking visual evidence of the *epistemic divide* through the lens of experimental practice. While initially sharing a common methodological ground dominated by *conventional lab experiments*, a significant divergence between the two clusters emerges and widens in the subsequent periods. The *Social Correlates of Trust* cluster progressively abandons the traditional lab setting, a shift culminating in the predominance of *framed field experiments* in the most recent period. In contrast, the *Social Preferences and Trust* cluster continues to prioritize the causal precision afforded by the *conventional lab experiments*, which remains its dominant methodology.

The divergence runs even deeper than Figure 6 suggests, extending to the spirit in which the experiments are conducted. For instance, the majority of *framed field experiments* in the *Social Preferences* cluster are conducted online, using platforms such as Amazon Mechanical Turk, which preserve the abstract and decontextualized nature of the lab.[21] This approach is fundamentally different from the immersive *lab-in-the-field* studies prevalent in the *Social Correlates* cluster, which bring controlled experimental methods into real-world environments with non-student participants (Favereau and Nagatsu 2025). Within this experimental economics tradition, the focus on social variables is so pervasive that it shapes even their *conventional lab experiments*. In this tradition, student subjects are often recruited not merely for convenience, but as the very object of study to investigate differences based on social identity, as in experiments comparing results across different geographical areas (Holm and Danielson 2005; Willinger et al. 2003) or testing for discrimination by recruiting students with "distinct ethnic affiliations" (Bornhorst et al. 2010; Fershtman and Gneezy 2001).

The choice of experimental method is not arbitrary but flows directly from the core objectives of each approach. The behavioral economics approach, driven by its need to test and refine formal theories, requires the utmost causal precision. To confidently claim that a behavior is motivated by a specific mechanism and not some other confounding factor, an environment free of external noise is essential. Consequently, the controlled setting of the *conventional lab experiment* is the preferred method; for this purpose, its artificiality is a feature, not a flaw. This choice, however, forces researchers to constantly grapple with specific challenges: the need for large samples to achieve adequate statistical power for subtle effects; the persistent threat of demand effects from subjects aware of being studied, which is particularly pronounced in the controlled laboratory setting; limited contextual richness compared to real-world environments; and resource and logistical constraints, such as the need for specialized facilities and personnel to run the experiments.

In contrast, the experimental economics approach is fundamentally concerned with the *generalizability* of its measures. This concern arises from the recognition that laboratory subjects may not represent the broader population. This imperative pushes researchers toward methods that improve representativeness, such as *artefactual* and *framed field experiments*. This path, however, trades one set of problems for another. Researchers face significant logistical hurdles in participant recruitment, from gaining the trust of specific communities to the sheer cost of fieldwork. Moreover, the gain in realism comes with a loss of control, as the inherent heterogeneity

[21] Mechanical Turk (MTurk) is an online crowdsourcing platform operated by Amazon, which allows individuals and organizations to post tasks (Human Intelligence Tasks, or HITs) that can be completed remotely by a large pool of registered workers. Participants are typically compensated with small payments for each task. In research, MTurk is commonly used as an alternative to traditional laboratory recruitment, allowing researchers to access a large pool of participants quickly and at low cost.

of participants and environmental *noise* create statistical challenges that can obscure the very causal effects being investigated.

The choice between the laboratory and the field is therefore more than a technical decision; it is a declaration of scientific priority. For behavioral economists, the artificiality of the lab is a necessary feature that allows for the clean isolation of theoretical mechanisms. For experimental economists, that same artificiality is a critical flaw that threatens the *generalizability* of the results. This clash reveals that the two clusters are not just using different methods; they are pursuing fundamentally different kinds of scientific truth. At the heart of this difference lie competing conceptions of what makes an experiment truly valid, a concept we turn to now.

3.3.3. The Meanings of External Validity

In economics, questions of validity largely fall within the dichotomy of *internal validity* versus *external validity*, concepts derived from "The Campbellian methodological project" (Jimenez-Buedo and Miller 2010, 302; Jiménez-Buedo and Russo 2021, 9571; Heukelom 2011, 21). As its name suggests, these concepts have been defined and refined by Campbell and his co-authors since the late 1950s (Campbell 1957; Campbell and Stanley 1966, Cook and Campbell 1979; Shadish, Cook, and Campbell 2001). In their latest work, validity is defined as the approximate truth of an inference: "when we say something is valid, we make a judgment about the extent to which relevant evidence supports that inference as true or correct" (Shadish, Cook, and Campbell 2001, 34). Building on this general notion they distinguish, among other distinctions, between *internal validity*—"the validity of inferences about whether observed covariation between A (the presumed treatment) and B (the presumed outcome) reflects a causal relationship from A to B as those variables were manipulated or measured"—and *external validity*—"the validity of inferences about whether the cause-effect relationship holds over variation in persons, settings, treatment variables, and measurement variables" (Shadish, Cook, and Campbell 2001, 38).

The challenge of applying these formal concepts becomes particularly clear when assessing the results of the Trust Game, which highlights that Campbell and his co-authors' definition of external validity leaves important ambiguities unresolved. This ambiguity, in turn, leads researchers to use *external validity* as a catch-all term, blurring the important distinction that should be made between *artificiality*–the correspondence of results to real-world behavior–and *generalizability*–their applicability across different samples and contexts (Reichardt 2011, 2019).[22] Drawing on a historical-methodological approach, Nagatsu and Favereau (2020) argues that this conceptual split emerges from the coexistence of distinct historical strands of experimentation within economics. The Trust Game literature clearly exemplifies this duality: studies focused on *generalizability* typically compare the game's outcomes with survey data on trust, whereas those focused on *artificiality* assess whether behavior within the game predicts real-world actions outside the laboratory.

[22] It is worth noting that economists' persistent focus on the *internal/external validity* dichotomy represents a selective adoption of the Campbellian tradition. As early as 1979, that framework evolved into a fourfold typology that also incorporates *statistical validity*–concerning the reliability of the statistical inference–and *construct validity*–concerning the relevance of the experimental setup to the theoretical concept (Cook and Campbell 1979). Economists, however, have largely continued to rely on the simpler dichotomy, overlooking this broader conceptualization (Shadish, Cook, and Campbell 2001; Jiménez-Buedo and Russo 2021).

Since the late 1960s, there have been two approaches to measuring trust in social sciences: experiments and surveys.[23] The three most commonly used sources in this literature are the *General Social Survey* (GSS), the *World Values Survey* (WVS) and the *German Socio-Economic Panel* (SOEP) (Hergueux and Jacquemet 2015, 259). Comparing the estimated level of trust in a population via surveys to the results of the Trust Game on a small sample of that population is an inquiry into the potential *generalizability* of Trust Game outcomes (i.e., their breadth). Put more figuratively, this approach seeks to assess whether the Trust Game serves as an accurate *social thermometer*.[24] This focus on *generalizability* therefore perfectly aligns with the inductive logic of the experimental economics approach, which aims to deploy the Trust Game as a tool for discovering robust empirical patterns with clear policy relevance.

Alós-Ferrer and Farolfi (2019) conducted a thorough investigation of studies that have challenged the *generalizability* of the Trust Game (Ermisch et al. 2009, Glaeser et al. 2000, Lazzarini et al. 2004). The authors subsequently concluded that "survey measures appear to have only a weak (if any) relationship with behavior in the Trust Game" (Alós-Ferrer and Farolfi 2019, 8). While this result may raise doubts about the Trust Game's ability to truly serve as a *social thermometer*, it is important to put these doubts into perspective. Employing surveys to measure a general level of trust is not without theoretical limitations, and the validity of these results themselves may be questioned (De Aquino 2017, Miller and Mitamura 2003).

The second approach to *external validity* directly tackles the question of *artificiality* by linking participants' behavior in the game to their actions in real-world settings. This approach to validity is much closer to the behavioral economics tradition, for which the central challenge is not to generalize a result from a sample to a population, but to generalize a theory from the lab to the world—where the ultimate test is not statistical generalizability, but the real-world relevance of the causal mechanisms the theory proposes. A systematic review by Galizzi and Navarro-Martinez (2019) provides a comprehensive overview of this literature. Research on *artificiality* differs methodologically from *generalizability* studies: while the latter predominantly rely on a single standardized proxy (survey data), the former links the Trust Game to a vast and heterogeneous range of real-world outcomes. These predicted variables span economic, social, and pro-social domains, including financial decisions like loan defaults (Karlan 2005), household investments (Bouma, Bulte, and van Soest 2008), labor market outcomes (Barr and Serneels 2009), or philanthropic behavior like donations (Baran, Sapienza, and Zingales 2010). Consequently, while this body of work demonstrates the game's predictive power across many domains, its sheer diversity makes it challenging to offer a single, definitive assessment of the *artificiality* of the Trust Game results.

In summary, the *epistemic divide* observed within the trust game literature is not a superficial disagreement over tools but a coherent cascade of interconnected philosophical and practical commitments. It originates from a fundamental divergence in scientific purpose: the hypothetico-deductive goal of theory adjudication on one side, and the inductive goal of empirical measurement on the other. This initial choice dictates the preferred type of experiment, creating a trade-off between the causal precision of the laboratory and the real-world relevance. Ultimately, this leads to a conceptual fault line where the shared term "external validity" itself acts as a veil, masking two fundamentally different scientific concerns: one of *artificiality* and the other

[23] See Dechaux (2017) for a discussion of the use of psychological tests in economics, and Heukelom (2011) for an exploration of the methodological differences between experimental psychology, psychological tests, and experimental economics.
[24] Guala (2008) provides insight on the use of paradigmatic games as a *social thermometer*.

of *generalizability*. The *epistemic divide* is therefore a consistent, multi-layered structure rooted in two distinct visions of what economic science should aim to achieve.

Ultimately, this *epistemic divide* can be envisioned as a fork in a research path. As we have demonstrated, this divergence originates from a single point—a difference in scientific ambition—but the paths chosen have led to increasingly distinct landscapes. Each tradition has cultivated its own specialized methods, its own experimental terrain, and its own language for discussing validity. The critical point, however, is that these paths are not running in parallel; they are actively moving apart.

To quantify the evolution of these inter-cluster relationships, we measure the normalized association strength (q_{ij}), a metric designed to capture the specificity of the bond between two research communities. This approach goes beyond simple link density by placing the affinity between two clusters in the context of their overall position within the research network. It is calculated as:

$$q_{ij} = \frac{w_{ij}}{\sqrt{\sum_k w_{ik} \cdot \sum_k w_{kj}}}$$

In this equation, the term w_{ij} represents the mean bibliographic coupling strength, calculated across all possible pairs of articles between the source cluster i and the target cluster j. The denominator establishes a baseline for this interaction by taking the geometric mean of the total outgoing citation strength of cluster i ($\sum_k w_{ik}$) and the total incoming citation strength of cluster j ($\sum_k w_{kj}$).[25]

As illustrated in Figure 7, the steady decline in this metric between the *Social Preferences and Trust* and *Social Correlates of Trust* clusters confirms that researchers in each tradition are increasingly relying on distinct bodies of literature. This growing distance, already visible in the diverging choices of experimental methodologies and scientific priorities, now points toward a future where the epistemic divide could solidify into a complete fracture. The risk is the formation of two isolated ecosystems of knowledge, each internally coherent but mutually unintelligible, thereby preventing the cross-pollination of ideas necessary for scientific progress. This isolation is particularly detrimental for the study of a phenomenon like trust, which inherently blends the psychological and social factors that these diverging traditions aim to study separately.

[25] This normalization allows us to measure how specific the bond between two clusters is relative to their other connections. For instance, a high and stable link density (w_{ij}) becomes increasingly specific and structurally significant if one of the clusters simultaneously weakens its ties to other fields. The q_{ij} score captures this dynamic precisely: a rising score indicates that a particular intellectual bond is not just strong but is becoming a defining and increasingly exclusive feature of a cluster's identity. As such, it provides a powerful quantitative correlate for the phenomena of intellectual convergence and fusion that we observe in the alluvial diagram.

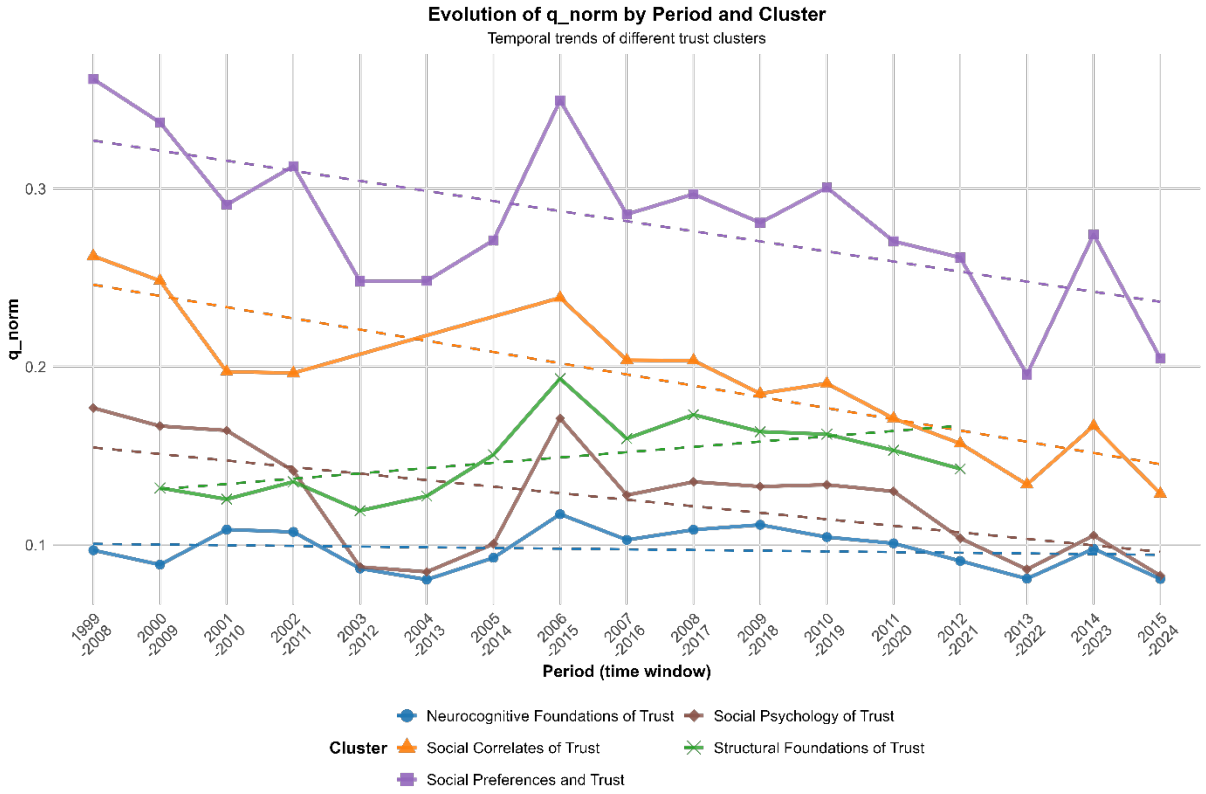


Figure 7 Temporal evolution of normalized interaction strength (q_norm) between "Social Preferences and Trust" and selected clusters. Higher q_norm values indicate stronger relative connections, reflecting how closely the cluster is linked to itself or to other clusters over time. Solid lines show observed values across time windows; dashed lines show linear trends.

4. Conclusion

This paper has traced the trajectory of the Trust Game, from its origins as a direct challenge to the paradigm of *calculative trust* to its current status as a canonical tool in the social sciences. The protocol conceived by Berg, Dickhaut, and McCabe in 1995 was a powerful experimental response to a paradigm that reduced trust to self-interest. By establishing a method to quantify varying levels of trust and trustworthiness in a non-zero-sum interaction, it offered a significant innovation over earlier frameworks. The initial findings, showing that significant trust could emerge even in a sterile, double-blind laboratory setting designed to encourage self-interested action, served as a compelling anomaly to rational choice theory and established trust as a legitimate, observable economic phenomenon. Its canonical status was then cemented by the robustness of its findings, as the core results withstood numerous stress tests and were ultimately confirmed on a massive scale by meta-analyses. It is this stability that transformed the game from an experimental curiosity into a powerful and versatile scientific instrument, thus shifting our historical narrative from one of conceptual development to one of widespread diffusion.

Our bibliometric analysis mapped this diffusion, revealing a scholarly landscape split along disciplinary and methodological lines, primarily between economics, psychology, and neuroscience. Within economics, this divide separates two distinct research programs. The first, aligned with behavioral economics, uses the Trust Game in a theory-inclined manner to build and test formal theories of social preferences. The second, rooted in experimental economics, employs the game inductively as a measurement tool to identify the social and demographic

correlates of trust, with a clear orientation toward informing public policy. As we have demonstrated, this is not a superficial disagreement over methods but rather a coherent, multilayered divide encompassing scientific approach, experimental design, and standards of validity. The growing separation between these two communities points toward a potential *epistemic fracture*, risking the formation of isolated and mutually unintelligible ecosystems of knowledge.

While this divide has produced increasingly separate research traditions, it does not necessarily doom the field to permanent fragmentation. Indeed, this divide also highlights a potential path toward a more unified economic approach to trust, one grounded in what Harrison and List (2004, 1010) call "methodological complementarity". In this perspective, the two traditions are not merely parallel; each seems to address what the other lacks. The theoretical depth of the behavioral tradition, with its reliance on controlled lab experiments, offers an ideal setting to tease apart potential confounds and isolate the precise *why* behind the robust empirical regularities uncovered by the experimental tradition. In turn, the latter's focus on real-world applicability can ground these abstract models in tangible outcomes, testing whether lab-validated mechanisms hold true across more complex social and economic environments.

Achieving this synthesis, however, faces significant hurdles. First, trust is an intrinsically complex phenomenon, combining psychological and social dimensions. Attempting to study these simultaneously creates significant confounding effects that are difficult to disentangle with the current tools of experimental and behavioral economics. This methodological challenge partly explains why the two traditions have specialized as they have. Beyond this intrinsic difficulty, these traditions are becoming more established on their distinct epistemic paths. Bridging this divide therefore demands more than a simple compromise; it requires a shared vision for a more comprehensive economic understanding of trust, one built on a concerted effort to design research programs where these distinct methods can "complement each other in some functional way" (Harrison and List 2004, 1009).

Acknowledgements

I am deeply grateful to Judith Favereau, Dorian Jullien, Ivan Moscati, and Alexandre Truc for their insightful comments on earlier versions of this paper. I also wish to thank the two anonymous reviewers, whose helpful suggestions significantly improved the quality and relevance of this article. All remaining errors and omissions are mine alone.

References

Ahlerup, Pelle, Ola Olsson, and David Yanagizawa. 2009. “Social Capital vs Institutions in the Growth Process.” European Journal of Political Economy 25 (1): 1–14. https://doi.org/10.1016/j.ejpoleco.2008.09.008.

Alós-Ferrer, Carlos, and Federica Farolfi. 2019. “Trust Games and Beyond.” Frontiers in Neuroscience 13: 887. https://doi.org/10.3389/fnins.2019.00887.

Angner, Erik. 2019. “We’re All Behavioral Economists Now.” Journal of Economic Methodology 26 (3): 195–207. https://doi.org/10.1080/1350178X.2019.1625210.

Apicella, Coren L., Anna Dreber, Benjamin Campbell, Peter B. Gray, Moshe Hoffman, and Anthony C. Little. 2008. “Testosterone and Financial Risk Preferences.” Evolution and Human Behavior 29 (6): 384–390. https://doi.org/10.1016/j.evolhumbehav.2008.07.001.

Aria, Massimo, and Corrado Cuccurullo. 2017. “Bibliometrix : An R-tool for Comprehensive Science Mapping Analysis.” Journal of Informetrics 11 (4): 959–975. https://doi.org/10.1016/j.joi.2017.08.007.

Ashraf, Nava, Iris Bohnet, and Nikita Piankov. 2006. “Decomposing Trust and Trustworthiness.” Experimental Economics 9 (3): 193–208. https://doi.org/10.1007/s10683-006-9122-4.

Balliet, Daniel, and Paul Van Lange. 2013. “Trust, Conflict, and Cooperation: A Meta-Analysis.” Psychological Bulletin 139 (5): 1090–1112. https://doi.org/10.1037/a0030939.

Banerjee, Ritwik. 2016. “Corruption, Norm Violation and Decay in Social Capital.” Journal of Public Economics 137: 14–27. https://doi.org/10.1016/j.jpubeco.2016.03.007.

Banerjee, Sanchayan, Matteo Galizzi, and Rafael Hortala-Vallve. 2021. “Trusting the Trust Game: An External Validity Analysis with a UK Representative Sample.” Games 12 (3): 66. https://doi.org/10.3390/g12030066.

Baran, Nicole, Paola Sapienza, and Luigi Zingales. 2010. “Can We Infer Social Preferences from the Lab? Evidence from the Trust Game.” SSRN Electronic Journal https://doi.org/10.2139/ssrn.1540137.

Barr, Abigail, and Pieter Serneels. 2009. “Reciprocity in the Workplace.” Experimental Economics 12 (1): 99–112. https://doi.org/10.1007/s10683-008-9202-8.

Berg, Joyce, John Dickhaut, and Kevin McCabe. 1995. “Trust, Reciprocity, and Social History.” Games and Economic Behavior 10 (1): 122–142. https://doi.org/10.1006/game.1995.1027.

Bohnet, Iris, Fiona Greig, Benedikt Herrmann, and Richard Zeckhauser. 2008. “Betrayal Aversion: Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States.” American Economic Review 98 (1): 294–310. https://doi.org/10.1257/aer.98.1.294.

Bohnet, Iris, and Richard Zeckhauser. 2004. “Trust, Risk and Betrayal.” Journal of Economic Behavior & Organization 55 (4): 467–484. https://doi.org/10.1016/j.jebo.2003.11.004.

Bolton, Gary E, and Axel Ockenfels. 2000. “ERC: A Theory of Equity, Reciprocity, and Competition.” The American Economic Review 90 (1): 166–193.

Bornhorst, Fabian, Andrea Ichino, Oliver Kirchkamp, Karl H. Schlag, and Eyal Winter. 2010. “Similarities and Differences When Building Trust: The Role of Cultures.” Experimental Economics 13 (3): 260–283. https://doi.org/10.1007/s10683-010-9240-x.

Bouma, Jetske, Erwin Bulte, and Daan van Soest. 2008. “Trust and Cooperation: Social Capital and Community Resource Management.” Journal of Environmental Economics and Management 56 (2): 155–166. https://doi.org/10.1016/j.jeem.2008.03.004.

Buchan, Nancy R., Rachel T. A. Croson, and Sara Solnick. 2008. “Trust and Gender: An Examination of Behavior and Beliefs in the Investment Game.” Journal of Economic Behavior & Organization 68 (3-4): 466–476. https://doi.org/10.1016/j.jebo.2007.10.006.

Camera, Gabriele, and Marco Casari. 2009. “Cooperation among Strangers under the Shadow of the Future.” American Economic Review 99 (3): 979–1005. https://doi.org/10.1257/aer.99.3.979.

Camerer, Colin. 2003. Behavioral Game Theory: Experiments in Strategic Interaction. The Roundtable Series in Behavioral Economics. New York, N.Y. : Princeton, N.J: Russell Sage Foundation ; Princeton University Press.

Camerer, Colin, and Richard Thaler. 1995. “Anomalies: Ultimatums, Dictators and Manners.” Journal of Economic Perspectives 9 (2): 209–219. https://doi.org/10.1257/jep.9.2.209.

Camilotto, Nicolas. 2021. “From Multidisciplinarity to Interdisciplinarity: on The Role of the Seminar as an Interface for Exchange, The case of the Seminar on Trust and Social change (1985-1986).” Revue d’économie politique 131 (5): 803–824. https://doi.org/10.3917/redp.315.0085.

Camilotto, Nicolas. 2023. “Navigating the Oceans of Research Literature on Trust.” OEconomia 13 (2): 201–237. https://doi.org/10.4000/oeconomia.15559.

Campbell, Donald. 1957. “Factors Relevant to the Validity of Experiments in Social Settings.” Psychological Bulletin 54 (4): 297–312. https://doi.org/10.1037/h0040950.

Campbell, Donald, and Julian Stanley. 1966. Experimental and Quasi-Experimental Designs for Research. Belomt, CA: Wadsworth.

Cassar, Alessandra, Pauline Grosjean, and Sam Whitt. 2013. “Legacies of Violence: Trust and Market Development.” Journal of Economic Growth 18 (3): 285–318.

Chai, Sun-Ki, Dolgorsuren Dorj, and Katerina Sherstyuk. 2018. “Cultural Values and Behavior in Dictator, Ultimatum, and Trust Games: An Experimental Study.” In Experimental Economics and Culture, Vol. 20, 0. Emerald Publishing Limited.

Charness, Gary, and Matthew Rabin. 2002. “Understanding Social Preferences with Simple Tests.” The Quarterly Journal of Economics 117 (3): 817–869.

Chaserant, Camille. 2008. “Les Fondements Incomplets de l’incomplétude : Une Revue Critique de La Théorie Des Contrats Incomplets*.” L’Actualité économique 83 (2): 227–253. https://doi.org/10.7202/017518ar.

Cherrier, Beatrice, and Andrej Svorenčík. 2018. “The Quantitative Turn in the History of Economics: Promises, Perils and Challenges.” Journal of Economic Methodology 25 (4): 367–377. https://doi.org/10.1080/1350178X.2018.1529217.

Chetty, Rinelle, Andre Hofmeyr, Harold Kincaid, and Brian Monroe. 2021. “The Trust Game Does Not (Only) Measure Trust: The Risk-Trust Confound Revisited.” Journal of Behavioral and Experimental Economics 90: 101520. https://doi.org/10.1016/j.socec.2020.101520.

Coleman, James. 1988. “Social Capital in the Creation of Human Capital.” American Journal of Sociology 94: S95–S120. https://doi.org/10.1086/228943.

Cook, Thomas, and Donald Campbell. 1979. Quasi-Experimentation: Design & Analysis Issues for Field Settings. Boston: Houghton Mifflin.

Dal Bó, Pedro, and Guillaume R. Fréchette. 2018. “On the Determinants of Cooperation in Infinitely Repeated Games: A Survey.” Journal of Economic Literature 56 (1): 60–114. https://doi.org/10.1257/jel.20160980.

Davis, John B. 2013. “Economics Imperialism under the Impact of Psychology: The Case of Behavioral Development Economics.” Œconomia. History, Methodology, Philosophy 119 (3-1): 138. https://doi.org/10.4000/oeconomia.638.

De Aquino, Jakson Alves. 2017. “The Meaning of Trust for Brazilians with Higher Education.” Social Indicators Research 130 (1): 325–349. https://doi.org/10.1007/s11205-015-1178-0.

DeBruine, Lisa M. 2002. “Facial Resemblance Enhances Trust.” Proceedings of the Royal Society of London. Series B: Biological Sciences 269 (1498): 1307–1312. https://doi.org/10.1098/rspb.2002.2034.

Dechaux, Pierrick. 2017. “L’économie Face Aux Enquêtes Psychologiques 1944-1960. Unité de La Science Économique, Diversité Des Pratiques.” PhD diss., Université Paris 1 Panthéon-Sorbonne.

Delgado, Mauricio, Hahn Frank, and Elizabeth Phelps. 2005. “Perceptions of Moral Character Modulate the Neural Systems of Reward during the Trust Game.” Nature Neuroscience 8 (11): 1611–1618. https://doi.org/10.1038/nn1575.

Desmet, Pieter T. M., David De Cremer, and Eric van Dijk. 2011. “Trust Recovery Following Voluntary or Forced Financial Compensations in the Trust Game: The Role of Trait Forgiveness.” Personality and Individual Differences 51 (3): 267–273. https://doi.org/10.1016/j.paid.2010.05.027.

Dufwenberg, Martin, and Georg Kirchsteiger. 2004. “A Theory of Sequential Reciprocity.” Games and Economic Behavior 47 (2): 268–298. https://doi.org/10.1016/j.geb.2003.06.003.

Dunning, David, Joanna E. Anderson, Thomas Schlösser, Daniel Ehlebracht, and Detlef Fetchenhauer. 2014. “Trust at Zero Acquaintance: More a Matter of Respect than Expectation of Reward.” Journal of Personality and Social Psychology 107 (1): 122–141. https://doi.org/10.1037/a0036673.

Edwards, José, Yann Giraud, and Christophe Schinckus. 2018. “A Quantitative Turn in the Historiography of Economics?” Journal of Economic Methodology 25 (4): 283–290. https://doi.org/10.1080/1350178X.2018.1529133.

Ellingsen, Tore, Magnus Johannesson, Sigve Tjøtta, and Gaute Torsvik. 2010. “Testing Guilt Aversion.” Games and Economic Behavior 68 (1): 95–107. https://doi.org/10.1016/j.geb.2009.04.021.

Engle-Warnick, J., and Robert L. Slonim. 2006. “Learning to Trust in Indefinitely Repeated Games.” Games and Economic Behavior 54 (1): 95–114. https://doi.org/10.1016/j.geb.2004.10.009.

Engle-Warnick, Jim, and Robert L. Slonim. 2004. “The Evolution of Strategies in a Repeated Trust Game.” Journal of Economic Behavior & Organization 55 (4): 553–573.

Ermisch, John, Diego Gambetta, Heather Laurie, Thomas Siedler, and Noah Uhrig. 2009. “Measuring People’s Trust.” Journal of the Royal Statistical Society Series A: Statistics in Society 172 (4): 749–769. https://doi.org/10.1111/j.1467-985X.2009.00591.x.

Ermisch, John F., and Diego Gambetta. 2006. “People’s Trust: The Design of a Survey-Based Experiment.” SSRN Electronic Journal https://doi.org/10.2139/ssrn.921059.

Favereau, Judith. 2021. Le hasard de la preuve: apports et limites de l’économie expérimentale du développement. Gouvernement en question(s). Lyon: ENS éditions.

Favereau, Judith, and Michiru Nagatsu. 2025. Field Experiments in Economics: History and Methodology. Routledge Studies in the History of Economics. Oxford: Taylor & Francis Routledge [Imprint].

Fehr, Ernst, and Klaus M. Schmidt. 1999. “A Theory of Fairness, Competition, and Cooperation.” The Quarterly Journal of Economics 114 (3): 817–868.

Fershtman, Chaim, and Uri Gneezy. 2001. “Discrimination in a Segmented Society: An Experimental Approach.” The Quarterly Journal of Economics 116 (1): 351–377.

Forsythe, Robert, Joel Horowitz, N.E Savin, and Martin Sefton. 1994. “Fairness in Simple Bargaining Experiments.” Games and Economic Behavior 6 (3): 347–369. https://doi.org/10.1006/game.1994.1021.

Foucault, Michel. 1989. Les mots et les choses: une archéologie des sciences humaines. Bibliothèque des sciences humaines. Paris: Gallimard.

Galizzi, Matteo, and Daniel Navarro-Martinez. 2019. “On the External Validity of Social Preference Games: A Systematic Lab-Field Study.” Management Science 65 (3): 976–1002. https://doi.org/10.1287/mnsc.2017.2908.

Glaeser, Edward, David Laibson, Jose Scheinkman, and Christine Soutter. 2000. “Measuring Trust.” Quarterly Journal of Economics 115 (3): 811–846. https://doi.org/10.1162/003355300554926.

Goutsmedt, Aurélien. 2021. “From the Stagflation to the Great Inflation: Explaining the US Economy of the 1970s.” Revue d’économie politique 131 (3): 557–582.

Goutsmedt, Aurélien, and Alexandre Truc. 2022. “Networkflow: Functions For A Workflow To Manipulate Networks.” .

Goutsmedt, Aurélien, Alexandre Truc, and François Claveau. 2021. “Biblionetwork: A Package For Creating Different Types of Bibliometric Networks. R Package.” .

Grossman, Sanford, and Oliver Hart. 1986. “The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration.” Journal of Political Economy 94 (4): 691–719. https://doi.org/10.1086/261404.

Grüne-Yanoff, Till. 2016. “Interdisciplinary Success without Integration.” European Journal for Philosophy of Science 6 (3): 343–360. https://doi.org/10.1007/s13194-016-0139-z.

Guala, Francesco. 2005. The Methodology of Experimental Economics. Cambridge ; New York: Cambridge University Press.

Guala, Francesco. 2008. “Paradigmatic Experiments: The Ultimatum Game from Testing to Measurement Device.” Philosophy of Science 75 (5): 658–669. https://doi.org/10.1086/594512.

Guala, Francesco. 2010. “Experimental Economics, History Of.” In Behavioural and Experimental Economics, edited by Steven N. Durlauf and Lawrence E. Blume, 99–106. London: Palgrave Macmillan UK.

Güth, Werner, Peter Ockenfels, and Markus Wendel. 1993. “Efficiency by Trust in Fairness? Multiperiod Ultimatum Bargaining Experiments with an Increasing Cake.” International Journal of Game Theory 22 (1): 51–73. https://doi.org/10.1007/BF01245570.

Güth, Werner, Rolf Schmittberger, and Bernd Schwarze. 1982. “An Experimental Analysis of Ultimatum Bargaining.” Journal of Economic Behavior & Organization 3 (4): 367–388. https://doi.org/10.1016/0167-2681(82)90011-7.

Hanson, Mark A., Pablo Gómez Barreiro, Paolo Crosetto, and Dan Brockington. 2024. “The Strain on Scientific Publishing.” Quantitative Science Studies 5 (4): 823–843. https://doi.org/10.1162/qss_a_00327.

Harrison, Glenn W., and John A. List. 2004. “Field Experiments.” Journal of Economic Literature 42 (4): 1009–1055. https://doi.org/10.1257/0022051043004577.

Hergueux, Jérôme, and Nicolas Jacquemet. 2015. “Social Preferences in the Online Laboratory: A Randomized Experiment.” Experimental Economics 18 (2): 252. https://doi.org/10.1007/s10683-014-9400-5.

Heukelom, Floris. 2011. “How Validity Travelled to Economic Experimenting.” Journal of Economic Methodology 18 (01): 13–28. https://doi.org/10.1080/1350178X.2011.556435.

Heyes, Anthony, and John List. 2016. “Supply and Demand for Discrimination: Strategic Revelation of Own Characteristics in a Trust Game.” American Economic Review 106 (5): 319–323. https://doi.org/10.1257/aer.p20161011.

Hirshleifer, Jack. 1977. “Economics from a Biological Viewpoint.” The Journal of Law and Economics 20 (1): 1–52. https://doi.org/10.1086/466891.

Holm, Håkan J., and Anders Danielson. 2005. “Tropic Trust Versus Nordic Trust: Experimental Evidence from Tanzania and Sweden.” The Economic Journal 115 (503): 505–532. https://doi.org/10.1111/j.1468-0297.2005.00998.x.

Isoni, Andrea, and Robert Sugden. 2019. “Reciprocity and the Paradox of Trust in psychological game theory.” Journal of Economic Behavior & Organization 167: 219–227. https://doi.org/10.1016/j.jebo.2018.04.015, Accessed 2025-10-25. https://www.sciencedirect.com/science/article/pii/S0167268118301215.

Jack, B. Kelsey. 2009. “Upstream–Downstream Transactions and Watershed Externalities: Experimental Evidence from Kenya.” Ecological Economics 68 (6): 1813–1824. https://doi.org/10.1016/j.ecolecon.2008.12.002.

Jimenez-Buedo, Maria, and Luis Miguel Miller. 2010. “Why a Trade-Off? The Relationship between the External and Internal Validity of Experiments.” THEORIA 25 (3): 301–321. https://doi.org/10.1387/theoria.779.

Jiménez-Buedo, María, and Federica Russo. 2021. “Experimental Practices and Objectivity in the Social Sciences: Re-Embedding Construct Validity in the Internal–External Validity Distinction.” Synthese 199 (3-4): 9549–9579. https://doi.org/10.1007/s11229-021-03215-3.

Johnson, Noel, and Alexandra Mislin. 2011. “Trust Games: A Meta-Analysis.” Journal of Economic Psychology 32 (5): 865–889. https://doi.org/10.1016/j.joep.2011.05.007.

Jullien, Dorian. 2018. “Under Risk, Over Time, Regarding Other People: Language and Rationality Within Three Dimensions.” Research in the History of Economic Thought and Methodology 36 (C): 119–155.

Kandori, Michihiro. 1992. “Social Norms and Community Enforcement.” The Review of Economic Studies 59 (1): 63–80. https://doi.org/10.2307/2297925.

Kaplan, Abraham. 1964. The Conduct of Inquiry: Methodology for Behavioral Science. 1st ed. Routledge.

Karlan, Dean. 2005. “Using Experimental Economics to Measure Social Capital and Predict Financial Decisions.” American Economic Review 95 (5): 1688–1699. https://doi.org/10.1257/000282805775014407.

Kessler, M. 1963. “Bibliographic Coupling between Scientific Papers.” American Documentation 14 (1): 10–25. https://doi.org/10.1002/asi.5090140103.

King-Casas, Brooks, Damon Tomlin, Cedric Anen, Colin Camerer, Steven Quartz, and Read Montague. 2005. “Getting to Know You: Reputation and Trust in a Two-Person Economic Exchange.” Science 308 (5718): 78–83. https://doi.org/10.1126/science.1108062.

Kosfeld, Michael, Markus Heinrichs, Paul Zak, Urs Fischbacher, and Ernst Fehr. 2005. “Oxytocin Increases Trust in Humans.” Nature 435 (7042): 673–676. https://doi.org/10.1038/nature03701.

Kreps, David. 1990. Corporate Culture And Economic Theory, 90–143. Cambridge University Press. Cambridge University Press.

Kreps, David M, and Robert Wilson. 1982. “Reputation and Imperfect Information.” Journal of Economic Theory 27 (2): 253–279. https://doi.org/10.1016/0022-0531(82)90030-8.

Laurent, Éloi. 2012. Économie de la confiance. Repères 595. Paris: la Découverte.

Lazzarini, Sergio, Regina Madalozzo, Rinaldo Artes, and José de Oliveira Siqueira. 2004. “Measuring Trust: An Experiment in Brazil.” Insper Working Paper 049 (2004).

Levi, Giovanni. 1992. “On Microhistory.” In New Perspectives on Historical Writing, edited by Peter Burke. University Park, Pa: Penn State University Press.

Lisciandra, Chiara. 2018. “The Role of Psychology in Behavioral Economics: The Case of Social Preferences.” Studies in History and Philosophy of Science Part A 72: 11–21. https://doi.org/10.1016/j.shpsa.2018.01.010.

Loewenstein, George. 1999. “Experimental Economics From the Vantage-point of Behavioural Economics.” The Economic Journal 109 (453): 25–34. https://doi.org/10.1111/1468-0297.00400.

Lorenz, Edward. 1996. “Confiance, contrats et coopération économique.” Sociologie du travail 38 (4): 487–508. https://doi.org/10.3406/sotra.1996.2282.

Lorenz, Edward. 1999. “Trust, Contract and Economic Cooperation.” Cambridge Journal of Economics 23 (3): 301–315. https://doi.org/10.1093/cje/23.3.301.

Magnússon, Sigurður Gylfi, and István M. Szijártó. 2013. What Is Microhistory ? Theory and Practice. London New York: Routledge.

McCabe, Kevin, and Vernon Smith. 2000. “A Comparison of Naïve and Sophisticated Subject Behavior with Game Theoretic Predictions.” Proceedings of the National Academy of Sciences 97 (7): 3777–3781. https://doi.org/10.1073/pnas.97.7.3777.

Miller, Alan, and Tomoko Mitamura. 2003. “Are Surveys on Trust Trustworthy?” Social Psychology Quarterly 66 (1): 62. https://doi.org/10.2307/3090141.

Nagatsu, Michiru, and Judith Favereau. 2020. “Two Strands of Field Experiments in Economics: A Historical-Methodological Analysis.” Philosophy of the Social Sciences 50 (1): 45–77. https://doi.org/10.1177/0048393119890393.

Nakayachi, Kazuya, and Motoki Watabe. 2005. “Influences of Hostage Posting on Estimation of Trustworthiness: The Effects of Voluntary Posting and Reliable Results.” Japanese Journal of Psychology 76 (3): 235–243. https://doi.org/10.4992/jjpsy.76.235.

Nowak, Martin A., and Karl Sigmund. 1998. “Evolution of Indirect Reciprocity by Image Scoring.” Nature 393 (6685): 573–577. https://doi.org/10.1038/31225.

Nowak, Martin A., and Karl Sigmund. 2005. “Evolution of Indirect Reciprocity.” Nature 437 (7063): 1291–1298. https://doi.org/10.1038/nature04131.

Ockenfels, Axel. 2010. The Selten School of Behavioral Economics: A Collection of Essays in Honor of Reinhard Selten. Berlin, Heidelberg: Springer Berlin / Heidelberg.

Ortmann, Andreas, John Fitzgerald, and Carl Boeing. 2000. “Trust, Reciprocity, and Social History: A Re-examination.” Experimental Economics 3 (1): 81–100. https://doi.org/10.1023/A:1009946125005.

Pillutla, Madan M., Deepak Malhotra, and J. Keith Murnighan. 2003. “Attributions of Trust and the Calculus of Reciprocity.” Journal of Experimental Social Psychology 39 (5): 448–455. https://doi.org/10.1016/S0022-1031(03)00015-5.

Rabin, Matthew. 1993. “Incorporating Fairness into Game Theory and Economics.” The American Economic Review 83 (5): 1281–1302.

Reichardt, Charles. 2019. Quasi-Experimentation: A Guide to Design and Analysis. Methodology in the Social Sciences. New York : London: The Guilford Press.

Reichardt, Charles S. 2011. “Criticisms of and an Alternative to the Shadish, Cook, and Campbell Validity Typology.” New Directions for Evaluation 2011 (130): 43–53. https://doi.org/10.1002/ev.364.

Revel, Jacques. 2010. “Microstoria.” In Historiographies: concepts et débats, edited by Christian Delacroix, François Dosse, Patrick Garcia, and Nicolas Offenstadt, Folio 179. Paris: Gallimard.

Roth, Alvin E. 1993. “The Early History of Experimental Economics.” Journal of the History of Economic Thought 15 (2): 184–209. https://doi.org/10.1017/S1053837200000936.

Rustichini, Aldo, and Marie Claire Villeval. 2014. “Moral Hypocrisy, Power and Social Preferences.” Journal of Economic Behavior & Organization 107: 10–24. https://doi.org/10.1016/j.jebo.2014.08.002.

Salton, Gerard, and Michael McGill. 1983. Introduction to Modern Information Retrieval. McGraw-Hill Computer Science Series. New York: McGraw-Hill.

Sapienza, Paola, Anna Toldra-Simats, and Luigi Zingales. 2013. “Understanding Trust.” The Economic Journal 123 (573): 1313–1332. https://doi.org/10.1111/ecoj.12036.

Scharlemann, Jörn P. W, Catherine C Eckel, Alex Kacelnik, and Rick K Wilson. 2001. “The Value of a Smile: Game Theory with a Human Face.” Journal of Economic Psychology 22 (5): 617–640. https://doi.org/10.1016/S0167-4870(01)00059-9.

Schechter, Laura. 2007. “Traditional Trust Measurement and the Risk Confound: An Experiment in Rural Paraguay.” Journal of Economic Behavior & Organization 62 (2): 272–292. https://doi.org/10.1016/j.jebo.2005.03.006.

Selten, Reinhard. 1989. “Evolution, Learning, and Economic Behavior.” Games and Economic Behavior 3 (1): 3–24. https://doi.org/10.1016/0899-8256(91)90003-W.

Selten, Reinhard. 2003. “Emergence and Future of Experimental Economics.” In Observation and Experiment in the Natural and Social Sciences, edited by Maria Carla Galavotti, 63–70. Dordrecht: Springer Netherlands.

Shadish, William, Thomas Cook, and Donald Campbell. 2001. Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Boston: Houghton Mifflin.

Slonim, Robert, and Pablo Guillen. 2010. “Gender selection discrimination: Evidence from a Trust game.” Journal of Economic Behavior & Organization 76 (2): 385–405. https://doi.org/10.1016/j.jebo.2010.06.016, Accessed 2025-10-25. https://www.sciencedirect.com/science/article/pii/S0167268110001368.

Smith, Vernon. 1982. “Microeconomic Systems as an Experimental Science.” The American Economic Review 72 (5): 923–955.

Smith, Vernon. 2020. “Trust, Reciprocity, and Social History: New Pathways of Learning When Max U (Own Reward) Fails Decisively.” ESI Working Papers .

Sugden, Robert. 2005. “Experiment, Theory, World: A Symposium on the Role of Experiments in Economics.” Journal of Economic Methodology 12 (2): 177–184. https://doi.org/10.1080/13501780500085943.

Svorenčík, Andrej. 2015. “The Experimental Turn in Economics: A History of Experimental Economics.” Rochester, NY, January.

Svorenčík, Andrej, and Harro Maas, eds. 2016. The Making of Experimental Economics: Witness Seminar on the Emergence of a Field. 1st ed., SpringerLink Bücher. Cham: Springer.

Tan, Jonathan H. W., and Claudia Vogel. 2008. “Religion and Trust: An Experimental Study.” Journal of Economic Psychology 29 (6): 832–848. https://doi.org/10.1016/j.joep.2008.03.002.

Tazdaït, Tarik. 2008. L'analyse économique de la confiance. 1st ed., Ouvertures économiques. Bruxelles: De Boeck.

Thielmann, Isabel, Giuliana Spadaro, and Daniel Balliet. 2020. "Personality and Prosocial Behavior: A Theoretical Framework and Meta-Analysis." Psychological Bulletin 146 (1): 30–90. https://doi.org/10.1037/bul0000217.

Trivers, Robert L. 1971. "The Evolution of Reciprocal Altruism." The Quarterly Review of Biology 46 (1): 35–57. https://doi.org/10.1086/406755.

Truc, Alexandre. 2022a. "Forty Years of Behavioral Economics." The European Journal of the History of Economic Thought 29 (3): 393–437. https://doi.org/10.1080/09672567.2021.1993295.

Truc, Alexandre. 2022b. "Interdisciplinary Influences in Behavioral Economics: A Bibliometric Analysis of Cross-Disciplinary Citations." Journal of Economic Methodology 29 (3): 217–251. https://doi.org/10.1080/1350178X.2021.2011374.

Truc, Alexandre, and Dorian Jullien. 2023. "A Controversy about Modeling Practices: The Case of Inequity Aversion." Journal of Economic Methodology 30 (3): 203–227. https://doi.org/10.1080/1350178X.2023.2180153.

Williamson, Oliver. 1993. "Calculativeness, Trust, and Economic Organization." The Journal of Law & Economics 36 (1): 453–486.

Willinger, Marc, Claudia Keser, Christopher Lohmann, and Jean-Claude Usunier. 2003. "A Comparison of Trust and Reciprocity between France and Germany: Experimental Investigation Based on the Investment Game." Journal of Economic Psychology 24 (4): 447–466. https://doi.org/10.1016/S0167-4870(02)00165-4.

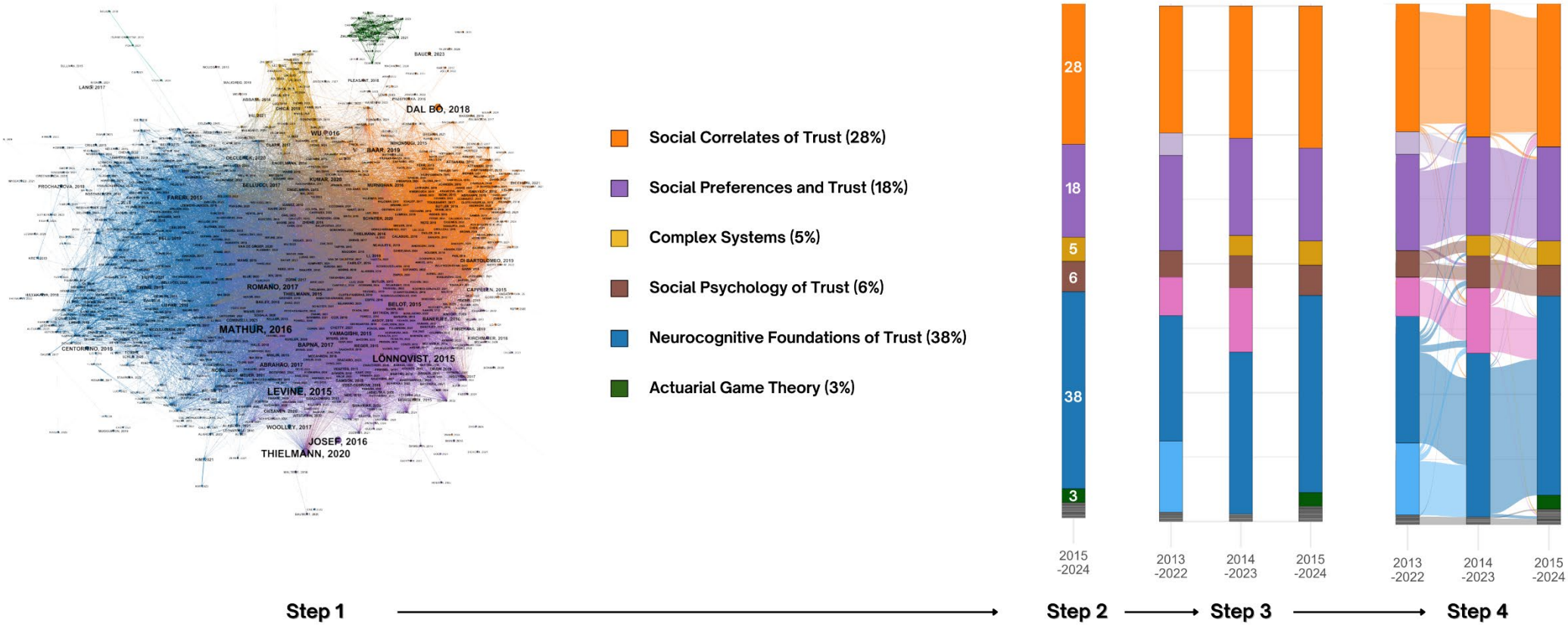
Wilson, Rick K., and Catherine C. Eckel. 2006. "Judging a Book by Its Cover: Beauty and Expectations in the Trust Game." Political Research Quarterly 59 (2): 189–202.

Wu, Junhui, Daniel Balliet, and Paul A. M. Van Lange. 2016. "Reputation Management: Why and How Gossip Enhances Generosity." Evolution and Human Behavior 37 (3): 193–201. https://doi.org/10.1016/j.evolhumbehav.2015.11.001.

Zak, Paul J., Robert Kurzban, and William T. Matzner. 2005. "Oxytocin Is Associated with Human Trustworthiness." Hormones and Behavior 48 (5): 522–527. https://doi.org/10.1016/j.yhbeh.2005.07.009.

Appendix

Appendix A. From Undirected Network Visualization to Alluvial Representation



Step 1 – Undirected network visualization: The bibliographic coupling network of the Trust Game literature (2015–2024) is visualized as a network of nodes and edges. Nodes represent publications and edges represent bibliographic coupling. The layout is computed using the ForceAtlas2 algorithm, which positions closely related publications near each other and spreads distant ones apart, making clusters of related research visually apparent.

Step 2 – Simplified cluster representation: To summarize the network, we represent the entire period (2015–2024) as a single bar. Within this bar, clusters are shown proportionally to their relative size in the network. This provides a compact overview of the distribution of research communities over the period.

Step 3 – Sliding the time window: The time window can be shifted, e.g., by one year, to observe how clusters evolve over time. By comparing consecutive periods, we can detect the growth, decline, or stability of individual clusters.

Step 4 – Inter-period relationships: Finally, we can visualize relationships between clusters across periods. This allows us to trace how publications and research communities transition from one period to the next, forming the basis for the construction of an alluvial diagram.

DOCUMENTS DE TRAVAIL GREDEG PARUS EN 2026
GREDEG Working Papers Released in 2026

2026-01 Estelle Malavolti & Frédéric Marty
L'équilibre économique des aéroports secondaires européens à l'épreuve des dynamiques de concurrence fiscale

2026-02 Charlie Joyez
A New Index of Export-Import Proximity: Conceptual Foundations and Global Patterns

2026-03 Patrice Bougette & Frédéric Marty
The SCP Paradigm Revisited: What Structuralism Really Contributed to U.S. Antitrust

2026-04 Imen Bouhlel, Nathalie Lazaric & Paolo Zeppini
Competitive Diffusion and Sustainability Transitions: The Case of Plastics Recycling Technologies

2026-05 Maxime Menuet
Fractional Replicator Dynamics

2026-06 Ilona Dielen, Patrice Bougette & Christophe Charlier
A (Green) Switch in Time Saves Nine: Assessing the Environmental Damage of the European Truck Cartel

2026-07 Thibault Schrepel & Godefroy de Boiscuillé
Questioning the Digital Markets Act's Legality

2026-08 Michael Finus & Paolo Zeppini
Green Lifestyles and Social Tipping Points

2026-09 Mathieu Chevrier & Sébastien Massoni
When Does Advisor Confidence Improve Decisions? Evidence from Human and Algorithmic Advice

2026-10 Flora Bellone, Charlie Joyez & Xavier Poulet-Goffard
Tax Avoidance by Small Multinationals as a Side Effect of Anti Tax Avoidance Policy

2026-11 Nicolas Camilotto
The Trust Game: A Historical and Methodological Analysis at the Frontier of Experimental and Behavioral Economics