THE COST OF INFORMATION:
THE CASE OF CONSTANT MARGINAL COSTS

By

Luciano Pomatto, Philipp Strack, and Omer Tamuz

COWLES FOUNDATION PAPER NO. 1854

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

2023

http://cowles.yale.edu/

Challenge

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/mcha20

John Maynard Keynes Narrates the Great Depression: His Reports to the Philips Electronics Firm

Robert W. Dimand & Bradley W. Bateman

To cite this article: Robert W. Dimand & Bradley W. Bateman (26 Oct 2023): John Maynard Keynes Narrates the Great Depression: His Reports to the Philips Electronics Firm, Challenge, DOI: 10.1080/05775132.2023.2272544

To link to this article: https://doi.org/10.1080/05775132.2023.2272544

Published online: 26 Oct 2023.

CHALLENGE
https://doi.org/10.1080/05775132.2023.2272544

John Maynard Keynes Narrates the Great Depression: His Reports to the Philips Electronics Firm†

Robert W. Dimand and Bradley W. Bateman

ABSTRACT
In October 1929, the Dutch electronics firm Philips approached John Maynatd Keynes to write confidential reports on the state of the British and world economies, which he did from January 1930 to November 1934, at first monthly and then quarterly. These substantial reports (Keynes's November 1931 report was twelve typed pages) show Keynes narrating the Great Depression in real time, as the world went through the US slowdown after the Wall Street crash, the Credit-Anstalt collapse in Austria, the German banking crisis (summer 1931), Britain's departure from the gold exchange standard in August and September 1931, the US banking crisis leading to the Bank Holiday of March 1933, the London Economic Conference of 1933, and the coming of the New Deal. This series of reports has not been discussed in the literature, though the reports and surrounding correspondence are in the Chadwyck-Healey microfilm edition of the Keynes Papers. We examine Keynes's account of the unfolding events of the early 1930s, his insistence that the crisis would be more severe and long-lasting than most observers predicted, and his changing position on whether monetary policy would be sufficient to promote recovery and relate his reading of contemporary events to his theoretical development.

Introduction

On October 23, 1929, just as Wall Street began to crash[1] and the world economy moved into exceptionally interesting times, Dr. H. F. van Walsem, counsel and secretary to the Dutch electronics firm N. V. Philips Gloeilampenfabrieken[2], wrote to "J. M. Keynes, Esq., C.B. Cambridge" asking him to write a monthly letter to the firm's Economic Intelligence Service about the state of the British economy and the world economy. John Maynard Keynes's letters to Philips, monthly from January 1930 to November 1931 and then, because of budget cuts to Philips's Economic Intelligence Service, quarterly from February 1932 to November 1934, show Keynes narrating the events of the Great Depression as they occurred, and reveal his perception of the convulsions of the

CONTACT Robert W. Dimand rdimand@brocku.ca, robert.dimand@yale.edu Department of Economics, Brock University, St. Catharines, Ontario L2S 3A1, Canada.

(Fall 2023: Visiting Professor of Economics, Yale University)
Robert W. Dimand is a professor of economics at Brock University, Canada, and a visiting professor at Yale University, and author of *The Origins of the Keynesian* and an editor of *The Elgar Companion to John Maynard Keynes*.
Bradley W. Bateman is president emeritus of Randolph College, Virginia, coauthor of *Capitalist Revolutionary: John Maynard Keynes* and an editor of *The Cambridge Companion to Keynes*.

world economy as he wrote his *General Theory of Employment, Interest and Money* (1936). This substantial body of Keynes's commentary on economic fluctuations (the November 1931 letter alone is twelve typed, double-spaced pages) has hitherto been neglected in the literature on Keynes. Keynes's reports and the associated correspondence, preserved in the Keynes Papers at King's College, Cambridge, are included in the 1993 Chadwyck-Healey microfilm edition of the Keynes Papers (section BM/5 Memoranda Exchanged with Business Houses), but the expense of this edition (which was sold only as a complete set of 170 reels of microfilm, priced at £9,700 or $17,000, plus $175 for a hardcover catalogue, Cox 1993) meant that only a few copies were sold. According to the WorldCat catalogue, there are five sets in libraries in the United States (Library of Congress, Harvard, Yale, Ohio State, and University of Texas at El Paso), two in Great Britain (Universities of Oxford and Sheffield), one in Canada (Victoria University in the University of Toronto) and a few in Germany (Göttingen), Italy and elsewhere but surprisingly little use has been made even of these copies of Keynes's letters to N. V. Philips. Neither Moggridge (1992) nor Skidelsky (1983–2000, 2003), major biographies of Keynes by the authors who know the Keynes Papers best, mentions Keynes's reports to Philips (but Backhouse and Bateman 2011, 129, have a paragraph about Keynes's July 1930 report). As Jacqueline Cox (1995, 171) notes, the thirty volumes of Keynes's *Collected Writings* (1971–1989) include "only a third of the bulk classified as economic" in the Keynes Papers at King's and do not include Keynes's philosophical papers there, while "the personal papers were barely touched." Donald Moggridge (2006, 136–137) observes that "There has, inevitably, been heavier use of the Keynes Papers in King's College Cambridge, which have the advantage of being available elsewhere on microfilm, than, say, his papers in the National Archives or his correspondence with his publishers, the last of which reveals the risks of depending on the Cambridge collection alone." A vast amount of research has been done about Keynes and his economics, yet not all the relevant material has been explored (see Backhouse and Bateman 2006, Dimand and Hagemann 2019).

These reports reveal Keynes's reading of what was happening in the British and world economies through the first four years of the Great Depression, and provide the empirical counterpart to the record of Keynes's theoretical development in this period given by notes taken by students at Keynes's lectures from 1932 to 1935 (Rymes 1987, 1989, Dimand 1988, Dimand and Hagemann 2019). After the success of *The Economic Consequences of the Peace* (1919), Keynes no longer needed to be paid for lecturing, and so gave a single series of eight lectures each year, on the subject of whatever book he was writing at the time, so his lectures from 1932 to 1935 are in effect annual drafts of the book that became *The General Theory*. These lectures at Cambridge and the reports to N. V. Philips on what was happening in the economy provide theoretical and empirical supplements to Keynes's *Collected Writings* (1971–1989), respectively, in following Keynes's intellectual development in the Great Depression, from *A Treatise on Money* (1930) to *The General Theory* (1936). In Keynes's workload, his reports to Philips from 1930 to 1934 took the place of the London and Cambridge Economics Service Special Memoranda on commodity markets that he wrote from 1923 to 1930 (Keynes [1923–30] 1983, 267–647), which provided an empirical counterpart to his normal backwardation theory of futures contracts ([1923] 1983, 1930, Chapter 29).

Replying on October 31 to von Walsem's letter inviting him to write the monthly letter to the firm's Economic Intelligence Service, Keynes was "quite ready to discuss this proposal with one of your representatives" but wished to clarify "that there will be no question of the publication of the letters and that they will be purely for the information of your own people" – and that "it would not be practicable to me to undertake such work except in return for a somewhat substantial fee which might be higher than you would be willing to offer." On November 4, von Walsem assured him that the letters would not be published and "There are only two persons who, though not in our service, are closely related to our firm, who also receive a copy of our Intelligence Service which they, however, are bound to consider as absolutely confidential." He suggested £100 a year. On November 13, Keynes, having "considered your kind proposal in relation to the fees which I have received on previous occasions for somewhat analogous work," offered to undertake the task for an initial six months, for £150 a year[3]. Although Van Walsem had initially asked for the suggestion of other authors if Keynes preferred not take on the task at the suggested £100 a year, and Keynes equally pointedly offered to suggest such alternative authors if Philips did not care to pay £150 a year, Van Walsem accepted Keynes's terms for Philips on November 22: "We think it desirable that one of our gentlemen will see you in order to discuss some details in the first half of December next."

In the event two representatives of Philips (Messrs. Sannes and du Pré) met with Keynes for a discussion summarized "for good order's sake" by van Walsem on December 21, 1929 (by which time van Walsem had already received a December 18 note by Keynes on the Australian exchange position). He recorded agreement that Keynes's monthly letter would treat "some important factor in the development of the British economic situation and give your opinion as to its effects on trade in general and on our business in particular. Also you will draw our attention to important events in the domains especially interesting us, in so far as these come to your knowledge … Whenever you think it necessary you will give us your views on the situation in different parts of the British Empire or eventually of other countries. If possible we shall suggest [to] you special points to be considered in your letters." Von Walsem wrote again on June 21, 1930 to confirm "that the arrangement has given us full satisfaction so that we are willing to continue on the same terms" and enclosed a cheque for 75 pounds. The arrangement also satisfied Keynes; he wrote on January 1, 1931, that "I have enjoyed preparing the letters." Keynes's letters balanced opinions about trade in general with observations about matters affecting Philips more specifically. Thus on January 11, 1930, Keynes stated that "The Factory capacity for Radio Sets seems to have become quite appalling during 1929" before proceeding more generally "to take this opportunity of emphasizing the anxiety which is felt here about the Australian position … I think that Australia may have more difficulties with her balance of trade during the coming year than the Argentine."[4]

The Slump of 1930: Investment, Debts and Deflation

Keynes's April 1930 letter suggested that, although a general improvement had not yet arrived, "there are a fair number of indications that we may be somewhere in the neighborhood of the bottom point." In particular, "the continuance of cheap money, and even more the expectation of such continuance, is bound to be effective in the situation in the course of a few months," but the effect on employment would be slower than on business feeling and the Stock Exchange and "it would not be surprising to see British unemployment figures go on mounting even to the neighborhood of 2,000,000 up to the end of this calendar year. … The effect of many rationalization schemes now in train will be for some time to come to improve profits rather than employment." With a large amount of Australian gold en route to the Bank of England, "there is less anxiety about the British exchange position than there has been for a very considerable time past" and Keynes expected the creation of the Bank for International Settlements to have a positive effect on confidence, a foreshadowing of his emphasis at Bretton Woods on the importance of designing appropriate international monetary institutions. Keynes doubted that the Federal Reserve Board would reverse its cheap money policy "until business and employment in the United States is a great deal better than it is now." This emphasis on expectations would be characteristic of Keynes's *General Theory* (although equally in line with Irving Fisher's quantity theoretic concern with expected inflation), as is the measurement of the ease of monetary policy by the cheapness of money, that is, by low nominal interest rates. Because nominal interest rates (especially short-term rates such as the Treasury Bill rate) were very low in a period of deflation, the Federal Reserve Board continued to view monetary conditions as easy throughout what Milton Friedman and Anna Schwartz (1963) later termed the "Great Contraction" of the US money supply (during which the monetary base increased, but not by enough to offset the rise in currency/deposit and reserve/deposit ratios), despite Fisher drawing the attention of his former student, Federal Reserve Governor Eugene Meyer, to the statistics on the shrinkage of the money supply, the sum of currency and demand deposits (Cargill 1992, Dimand 2019).

On June 24, 1930, H. du Pré emphasized that, "In reply to your remarks about the character of your monthly letters, we assure you that we leave it entirely to you to judge in each case which are the topics which are most worth being discussed by you." Nonetheless, "There is one question upon which we particularly should like to have your opinion." Keynes's monthly letters had repeatedly stated that recovery depended on the bond market becoming more active, with new loans being used not just for the refunding of floating debt but for new productive investment. "But on the other hand these last months many articles in the economic press" saw excessive capacity in many industries; "in other words that the world has first to grow into a productive apparatus which is too big for immediate needs. If this should be true, can a renewed investment-activity soon be hoped for, and if it soon comes, would it really do good? Of course there would be less unemployment in a number of industries; but would not prices of consumptive commodities, and so cost of living, rise? And especially it might turn out after some time, that the new activity has only added to the – supposed – actual over-investment, so that the disequilibrium would only be greater. It may of course be that entirely new industries are going to take the lead, but we do not yet see any that are very likely to do so. We should be much obliged if you would solve this puzzle for us or at least give your views on the pretended overcapacity and its probable effects on future developments in your next letter." This letter sheds light on the audience for Keynes's reports in the secretariat of N. V. Philips: not just salesmen looking for tips about the market for radio sets in Great Britain or elsewhere, but thoughtful businessmen pondering sophisticated economic issues such as the dual nature of productive investment in creating demand while increasing capacity (a problem to which the warranted growth rate of Harrod 1939 was an attempted solution).

In his July 1930 letter (seven typed pages, plus a six-page note on the bond market), Keynes warned that "it is now fully clear the world is in the middle of an international cyclical depression of unusual severity … a depression and a crisis of major dimensions … I believe that the prevailing opinion in the United States is still not pessimistic enough and is relying too much on a recovery in the early autumn, an event which is, in my opinion, most improbable. Nothing is more difficult than to predict the date of recovery. But all previous experience would show that a depression on this scale is not something from which the recovery comes suddenly or quickly." He felt that "The optimism of Wall Street and the hoarding tendencies of France may prevent any real recovery of the International Loan Market this year" and considered whether this might lead to "a psychological atmosphere in which really drastic scientific measures will be taken by Great Britain and the United States in conjunction to do what is humanly possible to cause a turn of the tide next spring. But one is traveling here into the realm of the altogether uncertain and unpredictable." In contrast, the Harvard Economic Society (founded by Harvard economics professors Charles J. Bullock and Warren Persons) stated in its weekly letter on June 28, 1930, that "irregular and conflicting movements of business should soon give way to sustained recovery" and on July 19 that "untoward elements have operated to delay recovery but the evidence nevertheless points to substantial improvement" (quoted by Galbraith 1961, 150, see also Walter Friedman 2014).

Responding to du Pré's query, Keynes reiterated that recovery would be preceded by "a substantial fall in the long-period rate of interest … leading in due course to the recovery of investment." But now he explained that he was not thinking of investment in manufacturing industry, "the world's capacity for which is probably quite ample for the present." Even at the highest estimate, the total cost of bringing Britain's industrial plant up to date "would not use up the country's savings for more than, say, three months. Moreover, when expected profits are satisfactory the rate of expenditure by manufacturing industry in fixed plant is not very sensitive to the rate of interest."

"On the other hand," in contrast to manufacturing, "the borrowing requirements for building, transport and public utilities are not only on a far greater scale, but are decidedly sensitive to the rate of interest. If I were to put my finger on the prime trouble to-day, I should call attention to the very high rate of interest for long-term borrowers … the long-term rate of interest is higher to-day than it has been in time of peace for a very long time past. When, at the same time, there is a big business depression and prices are falling, it is not surprising that new enterprise is kept back at the present level of interest." He drew attention to "those who might be called distress borrowers, that is say countries which have an urgent need for borrowing to pay off existing debts, and are consequently ready to pay a very high rate of interest," citing prospective Austrian, Hungarian and Australian loans on the London bond market, and remarked that "the effect of the German Loan has been to supply the French Treasury with funds, which it has withdrawn from the French market and is keeping unemployed in the

Bank of France." Keynes's July 1930 letter (discussed briefly by Backhouse and Bateman 2011, 129) illuminates both his analysis of the present situation and the role of investment in his economics. His distinction between investment in manufacturing, responsive to expected profit rather than interest rates, and interest-sensitive investment in construction, transport and public utilities clarifies his theory of investment. Increased investment was crucial for recovery of the world economy, and low long-term interest rates were necessary for high levels of investment in construction, transport and public utilities, the largest part of investment (even if manufacturing investment depended more on expected profits). In regard to the current situation, Keynes explained the forces getting long-term interest rates high even when prices were falling and short-term interest rates were low, but felt that "progress has been made toward getting the necessitous borrowers out of the way." On the immediate practical level, Keynes's distinction between the determinants of the two categories of investment dealt with du Pré's question of how low long-term interest rates could stimulate investment given excess productive capacity in manufacturing. And yet, unlike Harrod (1939), Keynes's July 1930 letter did not come to grips with the theoretical point raised by du Pré, the dual character of investment in creating both demand and productive capacity.

Keynes's August 1930 letter dissented from the view widely held in the United States "even in responsible quarters, that we may expect an autumn recovery with some confidence … a good deal of the American optimism is based on analogies drawn from the date of recovery after the 1920-21 slump" (compare the Harvard Economic Society's statement on August 30 that "the present depression has about spent its force," quoted by Galbraith 1961, 150). He argued that "Too much emphasis cannot be laid on the really catastrophic character of the price falls of some of the principal raw materials since a year ago" (even larger than appeared from published index numbers, because those included a number of commodities subject to price controls), which "must profoundly affect the purchasing power of all overseas markets." Long-term interest rates remained high, reducing new capital investment. In contrast, Keynes considered general opinion about the British position to be "perhaps a little too pessimistic." Britain was already in a difficult position before the slump of 1929 and 1930, because of the 1925 return to the gold exchange standard at the prewar parity (over the eloquent protests of Keynes 1925). But the heavy unemployment in the slump was limited to textiles and heavy industry (iron and steel, coal, and shipbuilding), export-based sectors already hit by the return to gold at an overvalued exchange rate (in his December 1930 letter, Keynes stated that if textiles, iron and steel, and coal were omitted, there was practically no decline in the Index of Production from a year before and an improvement from two years before). Keynes explained that British unemployment statistics, when used in international comparisons, "probably overstate the case" since the British statistics included "a great many workers in definite employment, but working short time … It is even the case that workers taking their normal summer holidays are now included in the figures of the unemployed." According to *The Economist*, the aggregate profits of all British joint stock companies reporting their earnings in the first half of 1930 "were not only greater than in the previous year, but were larger than in any previous year. This was partly due to the prosperity of British Oil Companies operating abroad, but by no means wholly." Nor did Keynes share the worries of financial opinion in London (and so some extent his own previous letter to Philips) about "the constant dribble of gold to France."

In Keynes's September 1930 letter to Philips, he was "still of the opinion that real recovery is a long way off. But at the same time it seems to me not unlikely that we are at, or near, the lowest point … It is time, therefore, to cease to be a 'bear', even if it is not yet time to be a 'bull'." His February 1931 letter began, "Glancing through the letters of previous months, I find that they were all extremely pessimistic (with a brief lapse into modified optimism in September, corrected in October). Nevertheless, in the light of the actual course of events they were scarcely pessimistic enough. Nor do I see any reason for expecting any appreciable alleviation in the coming months." His September 1930 letter reported that "An extraordinary example of the way in which a situation can suddenly turn round, when a tendency has been greatly overdone, has been seen on the London Stock Exchange in the last two weeks. There has been no recovery of business in Great Britain to account for it. The real facts are much as they were a month ago. But market pessimism, aided by bear operations, had brought security prices down to an absurdly low level not justified by the circumstances … everyone knew in his heart that prices were falling to foolish levels. The result was that within a few days the prices of many leading securities had risen from 10 to 20 per cent." The stock market had diverged from any level that could be construed as reflecting underlying fundamentals, but then abruptly bounced back. Keynes again stressed that Britain was not doing as badly as the United States in the slump: the fall in the British index of production from the previous year "is certainly less than 10 per cent" whereas the US index of industrial production for July 1930 was 37% below that for July 1929.

Keynes's 1930 "October Letter" warned that, "The catastrophic increase in the value of money has raised the burden of indebtedness of many countries beyond what they can bear … in many parts of the world the fall of prices has now reached a point where it is straining the social system at its foundations. Agriculturists and other producers of primary materials are being threatened with ruin and bankruptcy all over the world. It is useless to expect a recovery of markets in such conditions" (and in his February 1931 letter he again warned that "The prospect of a long series of defaults [by debtor countries exporting raw materials] during 1931 is not be excluded"). All of the gains that Germany had received in the Young Plan for reparations compared to the Dawes Plan were obliterated because "the clause in the Dawes Plan by which her [Germany's] liabilities in terms of gold were to be modified in the event of a change in prices was not included in the Young Plan." Keynes declared himself "rather more pessimistic … than a month ago." He remarked that in Britain, "Very slight steps have been taken, as yet, in the direction of reducing wages, which is probably inevitable, but will not get anyone much further if all countries alike embark on wage-cutting policies."

These themes of Keynes's October 1930 letter to Philips, the danger of ruin and bankruptcy from price deflation in a world where debts are fixed in money terms and the futility of wage-cutting, appeared publically in his December article in *The Nation and Atheneum* on "The Great Slump of 1930" (reprinted in his *Essays in Persuasion*, 1931). There Keynes (1931, 138–139) warned that, since wage and price deflation increases the real burden of debt and wage cuts reduce purchasing power, "neither the restriction of output nor the reduction of wages serves in itself to restore equilibrium" and went on to emphasize that "Moreover, even if we were to succeed eventually in reestablishing output at the lower level of money-wages appropriate to (say) the pre-war level of prices, our troubles would not be at an end. For since 1914 an immense burden of bonded debt, both national and international, has been contracted, which is fixed in terms of money. Thus every fall of prices increases the value of the money in which it is fixed. For example, if we were to settle down to the pre-war level of prices, the British National Debt would be nearly 40% greater than it was in 1924 and double what it was in 1920; … the obligations of such debtor countries as those of South America and Australia would become insupportable without a reduction of their standard of life for the benefit of their creditors; agriculturalists and householders throughout the world, who have borrowed o mortgage, would find themselves the victims of their creditors. In such a situation it must be doubtful whether the necessary adjustments could be made in time to prevent a series of bankruptcies, defaults, and repudiations which would shake the capitalist order to its foundations" (see also Dimand 2011). Here, before Fisher (1932, 1933, see Dimand 2019), was the concern with the effect of deflation on the real value of nominal deflation that reappeared in Chapter 19, "Changes in Money Wages," of *The General Theory*, where Keynes (1936, 264) warned that "if the fall of wages and prices goes far, the embarrassment of those entrepreneurs who are heavily indebted may soon reach the point of insolvency – with severely adverse effects on investment."

Contested Budgets, Trade Balance and the Banking and Exchange Crises of 1931

In 1930, Keynes's "November Letter" argued that foreign opinion underestimated the financial strength that accompanied Britain's industrial weakness: "it is forgotten that the adverse tendencies of the foreign exchanges, until recently, have been due, not to the absence of a favorable foreign trade balance, but to the eagerness of British investors to take advantage of the high profits or high rates of interest obtainable abroad. In 1929 the British favorable balance available for new foreign investment was greater than that for any other country, greater even than that for the United States. The Bank of England's difficulties were due to the fact that the pressure of savers to take advantage of opportunities abroad was even greater." Subsequent events in Wall Street and elsewhere had made overseas investment less appealing to British savers, so that the Bank of England was holding twenty million pounds sterling more of gold than a year before. In his December 1930 letter, Keynes reported that, even though "The perpetual drain of gold to France provides a source of nervousness and irritation in the money market" and although thirty million pounds sterling of gold had moved from Britain to France in the previous three months, the Bank of England held twenty-two million pounds sterling more in gold than a year before (but Keynes's March 1931 letter reported that a drain of twenty million pounds sterling of gold from the Bank of England in the previous three months "causing nervous talk to prevail in London"). Despite Keynes's repeated insistence on the financial strength of sterling and the growing gold reserves of the Bank of England (less than a year before the crisis of August and September 1931 that forced Britain off the gold exchange standard), the underlying message was that capital mobility under fixed exchange rates would constrain even the Bank of England from trying to lower long-term interest rates to stimulate investment. Until Britain left the gold standard and allowed sterling to float, Keynes's letters to Philips monitored the strength of protectionist sentiment in the British Government, but he lost interest in tariff proposals once the exchange rate was no longer pegged (see Keynes 1931). But there was one bright spot for Britain: Keynes's February 1931 letter stressed that "It must not be overlooked that England is gaining enormously by the tremendous drop in the price of her imports as compared with that of her exports."

Keynes's April 1931 letter to Philips is notable for explaining that Britain's apparent budget deficit of £23.5 million for the fiscal year ending March 31 "is not as bad as it sounds, since this figure is reached after allowing for the repayment of £67,000,000 of debt. So that, apart from debt repayments, there was a surplus on the year's workings of £43,500,000. It must be doubtful whether any other country is showing so favorable a result. Even if the sum borrowed for the unemployment fund, which lies outside the budget[5], were to be deducted, there would still have on the year a net reduction of debt." The next year's was expected to be larger, but "If no debt were to be repaid, there would probably be no deficit, even for the forthcoming year." Keynes's May 1931 letter, reporting on the budget presented by Labor Chancellor of the Exchequer Phillip Snowden, noted that "there will still be some reduction of debt during the forthcoming year, though not on as large as a scale as formerly." A few months later, when Snowden and Prime Minister Ramsay MacDonald broke with their party to join the Conservatives in a National Government to deal with a budget and exchange crisis, Snowden found it convenient to overlook that the apparent budget deficit was an artifact of budgeting for a reduction in the national debt, and to denounce his former Labor Cabinet colleagues for endangering the savings of small depositors by having the Post Office Savings Bank lend to the Unemployment Insurance Fund, without mentioned that such loans were guaranteed by the Treasury or that he had neglected to inform his Cabinet colleagues of the borrowing (as Keynes indignantly explained in two paragraphs in the draft of his November 1931 letter, deleted from the final version).

Keynes's May 1931 letter is also notable, in light of the subsequent exchange crisis that forced Britain off gold in September, for insisting that "The improvement in the sterling exchanges and the better gold position of the Bank of England, as it appears in the public returns, are not deceptive and may be assessed at even more than their face value." He held that "When there is no longer serious pressure on the Bank of England's gold, the stage will be set for really cheap money throughout the world … It will not mean a recovery, but it will pave the way for the recovery of investment which must precede the recovery of prices and profits." Keynes again emphasized that "the fall in the prices of the commodities imported by Great Britain has been so much greater than the fall in the prices of her exports. On the visible trade balance Great Britain was £5,000,000 better off in the first quarter of 1931 than in either of the preceding years … Thus the main burden of the present crisis falls on the raw-material-producing countries, and Great Britain is likely to gain gold in spite of the immense decline of her exports."

By the next month, as the Credit-Anstalt collapsed in Vienna (see Schubert 1991), as French and American capital then took flight from Germany (see Balderston 1994), and as share prices slumped in London, Wall Street and on most European bourses, Keynes felt "that we are now entering the crisis, or panic, phase of the slump. I am inclined to think that we look back on this particular slump we shall feel that this phase has been reached in the summer months of 1931, rather than at any earlier date." He warned that "the consequences of a change in the value of money, as reflected in the prices of leading commodities, so violent as that which has occurred in the last eighteen months, cannot be regarded too gravely. Until prices show a material rise the whole fabric of economic society will be shaken. Each decline of commodity prices and each further collapse on the Stock Exchanges of the world brings a further group of individuals or institutions into a position where their assets doubtfully exceed their liabilities."

Looking across the Atlantic: The American Slump

Keynes's July 1931 letter focused on the United States, where 21% of the industrial population was unemployed with perhaps another 20% working only two or three days a week: "it is quite out of the question that there should be anything which could be called a true recovery of trade at any time within, say, the next nine months. The necessary foundations for such a recover simply do not exist." Many of the loans of small banks to farmers or secured by real estate "are non-liquid and probably impaired. Thus there is a strong desire for the utmost liquidity while obtainable on the part of the ordinary Bank; and general unwillingness to take any unnecessary risks or to embark on speculative enterprise, even where the risk may be actuarially a sound one. The nervousness on the part of the Bankers is accompanied by a nervousness of the part of their depositors … So there is quite a common tendency to withdraw money from the banks and keep resources hoarded in actual cash … It was estimated that in the country as a whole as much as $500,000,000 was hoarded in actual cash in this way" (see Fisher 1933, Friedman and Schwartz 1963, Bernanke 2000). Keynes stressed that, "The American financial structure is more able than the financial structure of the European countries to support the strain of so great a change in the value of money. The very great development of Bank deposit and of bondage indebtedness in the United States means that a money contract has been interposed between the real estate on the one hand and the ultimate owner of the wealth on the other. The depreciation in the money value of the real estate sufficient to cause margins to run off, necessarily tends therefore to threaten the solidity of the structure."

Keynes reported in his July 1931 letter that although US agricultural wages had fallen by 20 to 25%, and there had also been large cuts to wages in small-scale industrial enterprises, hourly wages were practically unchanged for two thirds of the workers in large-scale industrial enterprises while the hourly wages of the other third had been reduced by some 10%. In October 1934, however, Keynes stated in his Cambridge lectures that "Labor will and has accepted reductions in money wages, in the USA in 1932, and it will not serve to reduce unemployment" with one student's notes calling the money-wage reductions "catastrophic" (Rymes 1987, 131).

Germany Defaults, Britain Abandons the Gold Parity

Turning from the United States, Keynes remarked near the end of his July letter that, "At the moment of writing there are heavy gold drains from London; but I do not think that this need be regarded with any undue alarm," a judgment that proved too sanguine.

More presciently, he added "The real danger in the situation comes from the possibility of the declaration of a general moratorium in Germany and the collapse of the mark [Germany defaulted on July 15]. The repercussion of such events on the solvency of the banking and money market systems of the world would be most serious." The next month, in his August 1931 newsletter (dated August 4), Keynes reported that "the bulk of the remaining short-term German debt is due to British and American banks and accepting houses; many accepting houses being landed with what are certainly frozen and may prove doubtful debts. Their own credit has suffered with the inevitable result, since they were the holders of large foreign balances, of a drain of gold from London … it would seem to be only ordinary prudence to act on the assumption that, while worse developments in Germany are doubtless possible, even apart from this the general underlying position is worse than the ordinary reader of newspapers believes it to be." While "Great Britain is suffering from the temporary shock to confidence due to the difficulties of the accepting houses,"[6] the situation of the world economy as a whole was more serious: "We are certainly standing in the midst of the greatest economic crisis of the modern world. Important though the German developments have been I would emphasize that these have been essentially consequences of deeper causes which are affecting all countries alike … For there is no financial structure which can withstand the strain of so violent a disturbance of values." A handwritten postscript at the end of the typed August 1931 letter warns Keynes's readers "not to be encouraged even by the appearance of apparently good news. The world financial structure is shaken and is rotten in many directions. Patching arrangements will be attempted, but they will not do much good, and it would be a mistake to place reliance on them." The next day, August 5, Keynes, writing to Prime Minister J. Ramsay MacDonald to urge rejection of the May Report, stated that "it is now virtually *certain* that we shall go off the existing parity at no distant date … when doubts, as to the prosperity of a currency, such as now exist about sterling, have come into existence, the game's up" (Keynes 1971–1989, Vol. XX, 591–593; Skidelsky 2003, 446), but he did not say so in print or to Philips – and he rejected, on patriotic grounds, a suggestion by O. T. Falk that the Independent Investment Trust, of which Keynes and Falk were directors, should replace a dollar loan with a sterling loan, which Keynes condemned as "a frank bear speculation against sterling." The Independent Investment Trust lost £40,000 by not switching its financing (Keynes 1971–1989, Vol. XX, 611–612; Moggridge 1992, 528–529; Skidelsky 2003, 447).

It was not only the world financial structure that was shaken; so was the Secretary Department of N. V. Philips. On August 6, 1931, H. du Pré wrote plaintively to Keynes, "Though we could hardly expect otherwise from your former letters, we note that you are not at all optimistic about the developments in the latter part of this year. These last weeks we read in the papers some statements from several Americans (among them people of authority), which hold a somewhat more cheerful view for the coming months. Must we infer from your letter that they are still, or again, too optimistic or is it possible that since your return from America[7] there have been some improvements, which may lead one to expect some improvement at least for the autumn?" Even Roger Babson, who had made his reputation by being bearish about the stock market in September 1929 (as he had been since 1926), was bullish by early 1931 (see W. Friedman 2014).

Keynes's reply on August 12 crushed any hopes: "In response to your enquiry, nothing has happened to make me more optimistic. As regards America, I consider that recovery this autumn is altogether out of the question. But the minds of all of us are of course dominated by the European and indeed the world situation. This still seems to me to be, as I have already described it, more serious than the general public know. I should recommend as complete inaction as is possible until further crises, or further striking events of some kind or another have occurred to clear up the situation."

Keynes's September letter (dated September 10, 1931), after the Conservative-dominated National Government displaced Labor, warned that "the hysterical concentration on Budgeting economy, which has also spread to the curtailment of expenditure by Local Authorities is calculated to produce unfavorable developments. For the widespread curtailment of expenditure is certain to reduce business profits and increase unemployment and lower the receipts of the Treasury, whilst it will do very little to tackle what is the fundamental problem, namely the improvement of the British Trade Balance. We seem likely to be faced by a period during which the balance of trade will not be sufficient to give confidence to foreign depositors."

It turned out, however, that one part of the cuts in government spending, the reduction in pay of the armed services, did indirectly dispose of the balance of payments problem. Since the government's version of equal sacrifice was that a vice-admiral earning £5 10s a day would lose 10 shillings a day (a reduction of 1/11), while naval lieutenants earning £1 7s a day and able-bodied seamen earning 5 shillings a day should each lose a shilling a day, reductions of 1/27 and 1/5, respectively (Muggeridge 1940, 109n), a naval mutiny erupted at Invergordon on September 16 (the first British naval mutiny since 1797), leading to abandonment of a fixed exchange rate on September 21 and a prompt 20% depreciation of sterling. Once the gold parity was abandoned, interest rates could be lowered without any balance of payments crisis. Commander Stephen King-Hall remarked "the strange combination of circumstances which caused the Royal Navy to be used by a far-seeing Providence as the unconscious means of … releasing the nation from the onerous terms of the contract of 1925 when the pound was restored to gold at pre-war parity … In 1805 the Navy saved the nation at Trafalgar; it may be that at Invergordon it achieved a like feat" (quoted by Muggeridge 1940, 111n). As for the budget deficit, Chancellor Snowden, who in the preceding Labor government had steadfastly blocked any reduction in the Sinking Fund contributions for paying down the national debt, now presented a budget reducing the annual Sinking Fund contribution by £20 million. Keynes declared in his October 1931 letter to Philips, "Great Britain's inevitable departure from the gold standard having occurred, it has been received with almost universal relief and in industrial circles a spirit of optimism is now abroad … Since the City and the Bank of England did their utmost to avoid the change, they feel that honor is satisfied. In other quarters the effect is to relieve a tension which was becoming almost unbearable … I have no doubt at all as to the reality of the stimulus which British business has obtained." Fisher (1935), assembling data on twenty-nine countries, found that recovery began only once a country abandoned the gold parity and was able to pursue a looser monetary policy (see Dimand 2003).

Keynes concluded his October 1931 letter, “The general passion for liquidity is bringing the value of cash in terms of everything else to so high a level as to be very near breaking point. This does not apply to Great Britain since her crisis was a balance of payments crisis rather than a banking crisis strictly so called. Thus the possibility of a general European and American banking crisis is the main risk, the possibility of which has now to be borne in mind.” The US banking crisis culminated in the “Bank Holiday” of March 1933, while all the major German and Italian banks passed into government ownership.

On November 3, 1931, Dr. du Pré was “very sorry to say that the necessity for the strictest economy which makes itself felt in all departments of our concern at present, impels us to an important curtailment of the budget of our Economic Intelligence Service” which would now issue bulletins every three months, instead of monthly. He asked Keynes for quarterly letters for £50 per annum, instead of monthly letters for £150 per annum. Keynes replied on November 9 that he read the letter “without any great surprise. I had been rather hesitating in my mind as to whether it is worth while to continue the arrangement on the new basis. But on the whole I feel that I should not like to break the friendly relations which have arisen between us, merely because times are bad.” He accepted the offer[8], asking to be reminded when each quarterly report was due, and enclosed his November letter stating that Britain was “to a considerable extent getting the best of both worlds since broadly speaking the countries from which we buy our food and raw materials have followed us off gold, whilst our manufacturing competitors have remained on the old gold parity.”[9] He felt that Continental observers were mistaken to think that Britain would want to return to gold: “Foreigners always underestimate the slow infiltration of what I have sometimes called ‘inside opinion’, whilst ‘outside opinion’ remains ostensibly unchanged. Then quite suddenly what ‘inside opinion’ becomes ‘outside opinion’. Foreigners are quite taken by surprise, but the change is really one which had been long prepared. In the later months of the old gold standard there was a hardly a soul in this country who really believed in it. But it was considered that it was our duty for fairly obvious reasons to do everything we possibly could to keep where we were.”

Keynes’s May 1932 quarterly letter stressed that, “The most important development, if one is thinking not so much of the moment but of laying the foundations for future improvement, is to be found in the return to cheap money, which was interrupted by the financial crisis of last summer and the departure from gold. I am more and more convinced in the belief, which I have held for some time, that an ultra-cheap money phase in the principal financial centers is an indispensable preliminary to recovery … Nevertheless it would be imprudent to expect too much at any early date from the stimulus of cheap money. The courage of enterprise is now so completely broken, that the effect on prices of money however cheap will be very slow. I consider it likely, therefore, that the cheap money phase may be extremely prolonged and that it may proceed to unprecedented lengths before it produces its effect.” He concluded, “For the time being the world is marking time, – waiting for it does not quite know what. I emphasize again the fact that the position in Great Britain, and in some of her Dominions, is relatively good. But for the time being, I see no light anywhere else … It would certainly be much too soon to take any steps whatever to be ready for a possible revival.”

Looking across the Atlantic: Hope from the New Deal

Keynes's August 1932 memorandum was notable for its explanation of why US stock prices had risen sharply and why that need not signal an end to the industrial crisis: the financial crisis had driven down stock prices until "the securities of many famous and successful companies were standing at little more than the equivalent of the net cash and liquid resources owned by those companies … the assets in question would either be worth nothing as a result of the general breakdown of contract, or must, in any circumstances apart from that, be worth a very great deal more than their quotations. Consequently, it is logical and right that the fear of their being worth nothing having been brought to an end, there should be a rapid recovery of the quotations on a very striking scale. It does not need a termination of the industrial crisis, or even an expectation of its early cessation, in order to justify the new levels."

In his February 1933 memorandum, commenting on the likely futility of the projected World Economic Conference, Keynes recalled that "I have myself put forward more drastic proposals for an international fiduciary currency, which would be the legal equivalent of gold. If this were agreed to, the position would be so much eased that various other desirable measures would also become practicable. I do not despair of converting British opinion to such a plan, but I am told that continental opinion would be almost unanimously opposed it." Keynes had contemplated such proposals long before Bretton Woods.

Keynes's August 1933 memorandum (actually mailed July 20, before Keynes left for holidays) held that "My own view is that President's Roosevelt's programme is to be taken most seriously as a means not only of American, but of world recovery. He will suffer set-backs and no one can predict the end of the story. But it does seem fairly safe to say that his drastic policies have had the result of turning the tide in the direction of better security not only in the United States, but elsewhere … Perhaps in the end President Roosevelt will devalue the dollar in terms of gold by 30 or 40 per cent." His November 1933 memorandum regretted "the failure of the President during his first six months to act inflation as well as talk it. In actual fact Governmental loan expenditure in the United States up to the end of September was on quite a trifling scale" but since then it seemed to be increasing: "if during the next six months the President is at last successful in putting into circulation a large volume of loan expenditure, I should expect a correspondingly rapid improvement in the industrial prosperity of America. This, if it occurs, would have a great influence on the rest of the world and especially on Great Britain … it might pave the way for a rate of improvement sufficiently rapid to deserve the name of real recovery." Keynes's February 1934 memorandum reported that in the United States "everything is moving strongly upwards. This is to be largely attributed to the fact that Governmental loan expenditure is now at last occurring on a large scale … the disbursement by the American Treasury of new money against borrowing has reached or is approaching $50,000,000 weekly and should maintain this rate for a few months to come." In his August 1934 memorandum, having visited the United States since his May memorandum, he found there "a recession which is somewhat more than seasonal," aggravated since his visit by a "failure of the corn crop … so acute as to be little short of a national disaster" but the actual and prospective level of US Government loan-financed expenditure made him optimistic about prospects for the US economy in the autumn and winter. He also reported that "the view is generally held in Great Britain that the gold block countries – including Holland not less than the others – cannot permanently maintain their present parity with gold without a disaster. Now or later it seems to us certain that the necessity for devaluation will be admitted." The reports end with Keynes's November 1934 memorandum, with no correspondence in the Keynes Papers concerning the end of his relationship with the Philips firm.

Conclusion: The Message of Keynes's Reports to Philips

Keynes's letters to the Philips electronics firm reveal he perceived events in the British and world economies from the beginning of 1930 through November 1934, and provide pungent and insightful commentary. These reports high-light the importance to Keynes of cheap money as a stimulus to investment – he was not just concerned with fiscal policy as the means to recovery, however much he placed emphasis from 1933 onward on the loan-financed expenditure of the Roosevelt Administration in the US. Keynes's response to a query from du Pré is particularly interesting about Keynes's distinction between those investment expenditures that are sensitive to interest rates and those that are not. The reports stress a theme discussed more briefly in Keynes's 1931 Harris Foundation lectures in Chicago (in Wright, ed., 1931) and in Chapter 19 of *The General Theory*, and at greater length by Irving Fisher (1932, 1933) (and later by Hyman Minsky 1975): since debt are contracted in nominal terms, a rise in the purchasing power of money increases the risk of bankruptcy, repudiation and default – and it is not just actual defaults that are costly, but also the perception of increased riskiness. Keynes recognized the exceptional seriousness of the Depression, dissenting firmly from predictions of an early recovery, and he saw clearly how defending overvalued gold parities forced central banks to keep interest rates high, instead of pursuing ultra-cheap money to restore investment. This hitherto-neglected body of evidence allows one to watch the unfolding of the world economic crisis of the early 1930s through Keynes's eyes, extraordinary events as viewed and narrated by an extraordinary economist. At £12 10s per report (by no means a trivial sum at the time), N. V. Philips certainly got their money's worth.

Notes

1. "Thursday, October 24, is the first of the days which history – such as it is on the subject – identifies with the panic of 1929" (Galbraith 1961, 103–104), but already on Monday, October 21, Irving Fisher had characterized the fall in stock prices as just the "shaking out of the lunatic fringe" and on Tuesday, Charles Mitchell of the National City Bank declared that "the decline has gone too far" (Galbraith 1961, 102).
2. Philips Incandescent Lamp Works, later Philips Electronics, successor to a firm founded by Lion Philips (originally Presburg), maternal uncle of Karl Marx (Gabriel 2011, 44, 110, 291-93, 295, 299, 315, 334, 366). Although relations between uncle and nephew were "strained by politics" (Gabriel 2011, 291), Mary Gabriel (2011, 299) refers to Marx's "fund of last resort, his uncle … He had sold himself to this pragmatic businessman as a successful writer only temporarily short of cash." Gabriel (2011, 642) remarks that "Marx's dabbling in the stock market has been questioned by some scholars, who believe he may simply have wanted his uncle to believe he was engaged in 'capital' transactions, not *Capital*." After the death of Lion

Philips, his sons did not reply to Marx's letter asking for help with his daughter Laura's wedding (Gabriel 2011, 364). Anthony Sampson (1968, 95) reported that the firm's chairman Frits Philips was "a keen Moral Rearmer and a fervent anti-communist, embarrassed by the fact that his grandfather was a cousin of Karl Marx."

3. For a sense of what £150 a year might have meant to Keynes: Moggridge (1992, 508, 585) and Skidelsky (2003, 417–418, 519, 565) report that Keynes's net worth fluctuated from £44,000 at the end of 1927 to £7,815 at the end of 1929, then rising to over £506,222 at the end of 1936, dropping again to £181,244 at the end of 1938. The offer from Philips came at a particularly low point in his finances. According to Skidelsky (2003, 265) "investment, directorship and consultancy income" accounted for more than 70% of Keynes's income between 1923-24 and 1928-29 (including £1,000 a year as chairman of National Mutual Life Assurance), books and articles for another 20%, leaving no more than a tenth of income from such academic sources as teaching, examining, being secretary of the Royal Economic Society and editor of its journal, and being Bursar and a Fellow of King's College.
4. However, writing to Keynes on January 21, H. du Pré was moved "to remark that the latest figures from the Argentine which, according to the handwritten note at the bottom of your letter, you intended to enclose, were not received here, so that we cannot give you an opinion about their importance for us."
5. When the majority report of the May Committee on National Expenditure projected on July 31, 1931, that the budget deficit for 1931-32 would be £120 million, necessitating £96 million of cuts to unemployment benefits, road construction, and government and armed forces pay, it counted all borrowing by the Unemployment and Road funds as "public expenditure on current account" as well as "the usual provision for the redemption of debt" of £50 million (Winch 1969, 126–130). Keynes accused the majority on the May Committee of not "having given a moment's thought to the possible repercussions of their programme, either on the volume of unemployment or on the receipts of taxation" – he estimated it would add 250,000 to 400,000 to the unemployed, and reduce tax receipts by £70 million (*New Statesman and Nation*, August 15, 1931; Keynes 1971-89, Vol. IX, 141–145; Winch 1969, 130, Skidelsky 2003, 446).
6. With regard to Britain, Keynes noted that "There is, however, tremendous pressure of public opinion towards the Government Economy, which means in the main a reduction in the salaries of Government employees and of the allowances of the unemployed. It is equally difficult for the present [Labour] Government either to refuse or concede concessions to this trend of opinion. But if a movement in this direction takes place, which is still most doubtful, it remains exceedingly open to argument whether the result on the actual level of unemployment will be favourable."
7. Keynes had given three Harris Foundation Lectures on "An Economic Analysis of Unemployment" at the University of Chicago in June and July 1931, published in Quincy Wright, ed. (1931), and reprinted in Keynes (1971-89), Vol. XIII. These lectures mostly expounded the analysis of Keynes's *Treatise*, but the third lecture also examined the debt-deflation process, the undermining of the financial structure by an increase in the real value of debts and fall in the nominal value of collateral (Keynes 1971-89, Vol. XIII, 359–361, see Dimand 2011).
8. He also raised a "small personal matter", asking for advice on buying a new wireless set that would "have a thoroughly good loud speaker, both for voice and music reproduction and should be able to pick up distant stations such as Moscow."
9. A passage crossed-out in the draft of Keynes's November 1931 letter, in the section discussing the general election, stated that, "As has been the case in the last three or four General Elections, it is that old wretch Lord Rothermere [publisher of the *Daily Mail*] who has been dead right. It is said that he has made a profit on the crisis of £100,000, buying majorities on the Stock Exchange." Skidelsky (2003, 472) relates that Keynes "consistently lost money (his own and his friends') on the results of general elections."

References

Backhouse, Roger E., and Bradley W. Bateman, eds. 2006. *The Cambridge Companion to Keynes*. Cambridge, UK: Cambridge University Press.

Backhouse, Roger E., and Bradley W. Bateman. 2011. *Capitalist Revolutionary: John Maynard Keynes*. Cambridge, MA: Harvard University Press.

Balderston, Theo. 1994. “The Banks and the Gold Standard in the German Financial Crisis of 1931.” *Financial History Review* 1 (1):43–68. https://doi.org/10.1017/S0968565000001554

Bernanke, Ben S. 2000. *Essays on the Great Depression*. Princeton, NJ: Princeton University Press.

Cargill, Thomas F. 1992. “Miscellany: Irving Fisher Comments on Benjamin Strong and the Federal Reserve in the 1930s.” *Journal of Political Economy* 100 (6):1273–7. https://doi.org/10.1086/261861

Chadwyck-Healey. 1993. *The John Maynard Keynes Papers in King’s College, Cambridge, 170 Reels of Microfilm*. Cambridge, UK: Chadwyck-Healey Ltd.

Cox, Jacqueline. 1993. *A Catalogue of the John Maynard Keynes Papers in King’s College, Cambridge*. Cambridge, UK: Chadwyck-Healey Ltd.

Cox, Jacqueline. 1995. “Keynes: An Archivist’s View.” In *New Perspectives on Keynes,* Annual Supplement to History of Political Economy. Vol. 27, edited by Allin F. Cottrell and Michael S. Lawlor, 163–75. Durham, NC: Duke University Press.

Dimand, Robert W. 1988. *The Origins of the Keynesian Revolution*. Aldershot, UK: Edward Elgar Publishing.

Dimand, Robert W. 2003. “Irving Fisher on the International Transmission of Booms and Depressions through Monetary Standards.” *Journal of Money, Credit and Banking* 35 (1):49–90. https://doi.org/10.1353/mcb.2003.0002

Dimand, Robert W. 2011. “The Consequences to the Banks of the Collapse of Money Values,’ 1931 and 2009.” In *Perspectives on Keynesian Economics*, edited by Arie Arnon, Jimmy Weinblatt, and Warren Young, 233–45. Heidelberg: Springer Verlag.

Dimand, Robert W. 2019. *Irving Fisher*. London: Palgrave Macmillan, Great Thinkers in Economics.

Dimand, Robert W., and Harald Hagemann, eds. 2019. *The Elgar Companion to John Maynard Keynes*. Cheltenham, UK.

Fisher, Irving. 1932. *Booms and Depressions*. New York: Adelphi.

Fisher, Irving. 1933. “The Debt-Deflation Theory of Great Depressions.” *Econometrica* 1 (4):337–57. https://doi.org/10.2307/1907327

Fisher, Irving. 1935. “Are Booms and Depressions Transmitted Internationally through Monetary Standards?” *Bulletin de L’Institut Internationale de Statistique* 28 (1):1–29. Reprinted in Dimand (2003).

Friedman, Milton, and Anna J. Schwartz. 1963. *A Monetary History of the United States, 1867–1960*. Princeton, NJ: Princeton University Press for the National Bureau of Economic Research.

Friedman, Walter A. 2014. *Fortune Tellers: The Story of America’s First Economic Forecasters*. Princeton, NJ: Princeton University Press.

Gabriel, Mary. 2011. *Love and Capital: Karl and Jenny Marx and the Birth of a Revolution*. New York: Little, Brown.

Galbraith, John Kenneth. 1961. *The Great Crash 1929, with New Introduction*. Boston, MA: Houghton Mifflin.

Harrod, Roy F. 1939. “An Essay in Dynamic Theory.” *The Economic Journal* 49 (193):14–33. https://doi.org/10.2307/2225181

Keynes, John Maynard. 1983 [1923]. “Some Aspects of Commodity Markets.” *Manchester Guardian Commercial, Reconstruction in Europe* 29 March 1923; reprinted in Keynes (1971–89), XII:255–66.

Keynes, John Maynard. 1983 [1923–30]. *London and Cambridge Economic Service Special Memoranda*, as reprinted in Keynes (1971–89), XII: 267–647.

Keynes, John Maynard. 1919. *The Economic Consequences of the Peace*. London: Macmillan.

Keynes, John Maynard. 1925. *The Economic Consequences of Mr. Churchill*. London: Leonard and Virginia Woolf, at the Hogarth Press.
Keynes, John Maynard. 1930. *A Treatise on Money*, 2 Volumes. London: Macmillan.
Keynes, John Maynard. 1931. *Essays in Persuasion*. London: Macmillan.
Keynes, John Maynard. 1936. *The General Theory of Employment, Interest and Money*. London: Macmillan.
Keynes, John Maynard. 1971–89. *The Collected Writings of John Maynard Keynes*, 30 volumes, general editors Donald E. Moggridge and E. A. G. Robinson, volume editors Elizabeth S. Johnson and Donald E. Moggridge. London: Cambridge University Press, for the Royal Economic Society.
Minsky, Hyman P. 1975. *John Maynard Keynes*. New York: Columbia University Press.
Moggridge, Donald E. 1992. *Maynard Keynes: An Economist's Biography*. London: Routledge.
Moggridge, Donald E. 2006. "Keynes and His Correspondence." In *The Cambridge Companion to Keynes*, edited by Roger E. Backhouse and Bradley W. Bateman, 136–59. Cambridge, UK: Cambridge University Press.
Muggeridge, Malcolm. 1940. *The Thirties: 1930-40 in Great Britain*. London: Hamish Hamilton.
Rymes, Thomas K., ed. 1987. *Keynes's Lectures: Notes of Students*. Ottawa: Department of Economics, Carleton University.
Rymes, Thomas K., ed. 1989. *Keynes's Lectures, 1932–35: Notes of a Representative Student*. London: University of Michigan Press.
Sampson, Anthony. 1968. *The New Europeans*. London: Hodder and Stoughton.
Schubert, A. 1991. *The Credit-Anstalt Crisis of 1931*. Cambridge, UK: Cambridge University Press.
Skidelsky, Robert. 1983–2000. *John Maynard Keynes*, 3 volumes. London: Viking.
Skidelsky, Robert. 2003. *John Maynard Keynes, 1883-1946: Economist, Philosopher, Statesman*. London: Macmillan.
Winch, Donald. 1969. *Economics and Policy: A Historical Study*. New York: Walker and Company.
Wright, Quincy, ed. 1931. *Unemployment as a World Problem: The Harris Foundation Lectures*. Chicago, IL: University of Chicago Press.

American Economic Review 2023, 113(5): 1360–1393
https://doi.org/10.1257/aer.20190185

The Cost of Information: The Case of Constant Marginal Costs†

By Luciano Pomatto, Philipp Strack, and Omer Tamuz*

We develop an axiomatic theory of information acquisition that captures the idea of constant marginal costs in information production: the cost of generating two independent signals is the sum of their costs, and generating a signal with probability half costs half its original cost. Together with Blackwell monotonicity and a continuity condition, these axioms determine the cost of a signal up to a vector of parameters. These parameters have a clear economic interpretation and determine the difficulty of distinguishing states. (*JEL* D82, D83)

Much of contemporary economic theory is built on the idea that information is scarce and valuable. A proper understanding of information as an economic commodity requires theories for its value, as well as for its production cost. While the literature on the value of information (Bohnenblust, Shapley, and Sherman 1949; Blackwell 1951) is by now well established, modeling the cost of producing information has remained an unsolved problem.[1] In this paper, we develop an axiomatic theory of costly information acquisition.

We characterize all cost functions over Blackwell experiments that satisfy three main axioms: First, experiments that are more informative in the sense of Blackwell (1951) are more costly. Second, the cost of generating independent experiments equals the sum of their individual costs. Third, the cost of generating an experiment with probability half equals half the cost of generating it with probability one.

Our three axioms admit a straightforward economic interpretation. The first one is a form of monotonicity: more precise information is more costly. The second and third axioms capture the idea of linear cost. The second axiom implies that the cost of collecting n independent random samples is linear in n. For example,

*Pomatto: Caltech (email: luciano@caltech.edu); Strack: Yale (email: philipp.strack@gmail.com); Tamuz: Caltech (email: tamuz@caltech.edu). Pietro Ortoleva was the coeditor for this article. Philipp Strack was supported by a Sloan Fellowship. Omer Tamuz was supported by a grant from the Simons Foundation (419427), a Sloan research fellowship, a BSF award (2018397), and an NSF CAREER award (DMS-1944153). We thank Kim Border, Ben Brooks, Simone Cerreia-Vioglio, Tommaso Denti, Federico Echenique, Drew Fudenberg, Ed Green, Adam Kapor, Massimo Marinacci, Jeffrey Mensch, Filip Matějka, Stephen Morris, Teemu Pekkarinen, Doron Ravid, and Yangfan Zhou for their comments. All errors and omissions are our own.

†Go to https://doi.org/10.1257/aer.20190185 to visit the article page for additional materials and author disclosure statement(s).

[1] For example, Arrow (1985, p. 303) makes the following statement: "The choice of information structures must be subject to some limits, otherwise, of course, each agent would simply observe the entire state of the world. There are costs of information, and it is an important and incompletely explored part of decision theory in general to formulate reasonable cost functions for information structures."

if the variable is the perceived quality of a new product, and information is generated by surveying random customers, the axiom is satisfied if the cost of calling an additional customer is constant: i.e., calling 20 customers is twice as costly as calling 10. More generally, the axiom requires the cost to be additive with respect to experiments that are independent conditional on the state. Similarly, the third axiom implies that the cost of producing a sample with probability α is a fraction α of the cost of acquiring the same sample with probability one. This axiom is satisfied by all posterior separable costs, which include nearly all models of information cost in the literature.

We propose these linearity assumptions as a way of studying cost functions over information structures. In the context of traditional commodities, a standard avenue for studying cost functions is by categorizing them in terms of decreasing, increasing, or constant marginal costs, with the latter being arguably the conceptually simplest case. In this paper we take a similar approach for studying the cost of information acquisition, and our axioms make an attempt at formalizing the assumption of constant marginal costs for information. As in the case of traditional commodities, assuming linear costs is restrictive, and it is easy to conceive of decision problems where our axioms are violated. For example, if customers are hard to find, surveying 20 customers might cost more than twice as much as surveying 10. Conversely, economies of scale may result in decreasing marginal costs. Nevertheless, our axioms have the advantage of admitting a clear economic interpretation, making it possible to judge for which applications they are appropriate. We thus propose the study of linear cost functions as a first step towards the wider goal of studying general information costs in terms of their economic properties.

Representation.—The main result of this paper is a characterization theorem for cost functions over experiments. We are given a finite set Θ of states of nature. An experiment μ produces a signal realization $s \in S$ with probability $\mu_i(s)$ in state $i \in \Theta$. We show that for any cost function C that satisfies the above postulates, together with a continuity assumption, there exist unique nonnegative coefficients (β_{ij}), one for each ordered pair of states of nature i and j, such that[2]

$$C(\mu) = \sum_{i,j\in\Theta} \beta_{ij}\left(\sum_{s\in S} \mu_i(s)\log \frac{\mu_i(s)}{\mu_j(s)}\right). \tag{1}$$

Each coefficient β_{ij} can be interpreted as capturing the difficulty of discriminating between state i and state j, as the cost can be expressed as a linear combination

$$C(\mu) = \sum_{i,j\in\Theta} \beta_{ij} D_{\mathrm{KL}}(\mu_i \,\|\, \mu_j),$$

where the *Kullback-Leibler divergence*

$$D_{\mathrm{KL}}(\mu_i \,\|\, \mu_j) = \sum_{s\in S} \mu_i(s)\log \frac{\mu_i(s)}{\mu_j(s)}$$

[2]Throughout the paper we assume that the set of states of nature Θ is finite. We do not assume a finite set S of signal realizations and the generalization of (1) to infinitely many signal realizations is given in (3).

is the expected log-likelihood ratio between state i and state j when the state equals i. The term $D_{\mathrm{KL}}(\mu_i \| \mu_j)$ is thus large if the experiment μ on average produces evidence that strongly favors state i over j, conditional on the state being i. Hence, the larger the coefficient β_{ij}, the more costly it is to reject the hypothesis that the state is j when it truly is i. Formally, β_{ij} is the marginal cost of increasing the expected log-likelihood ratio of an experiment with respect to states i and j, conditional on i being the true state. We refer to the cost (1) function as the *log-likelihood ratio cost* (or *LLR cost*).

In many common information acquisition problems, states of the world are one-dimensional quantities. For instance, this is the case when the unknown state is a physical quantity such as height or weight, or an economic quantity such as the inflation rate. In these examples, an experiment can be seen as a noisy measurement of the unknown underlying state $i \in \mathbb{R}$. We provide a framework for choosing the coefficients β_{ij} in these contexts. Our main hypotheses are that the difficulty of distinguishing between two states i and j is a function of the distance between them, and that the cost of performing a measurement with standard Gaussian noise does not depend on the set of states Θ in the particular information acquisition problem; this is a feature that is commonly assumed in models that exogenously restrict attention to normal experiments.

Under these assumptions, we show that there exists a constant $\kappa \geq 0$ such that, for every pair of states $i,j \in \Theta$,

$$\beta_{ij} = \frac{\kappa}{(i-j)^2}.$$

In this functional form, the difficulty of distinguishing between states is a quadratic decreasing function of the distance between them. As we show, this choice of parameters offers a simple and tractable framework for analyzing the implications of the LLR cost.

The concept of a Blackwell experiment makes no direct reference to subjective probabilities nor to Bayesian reasoning.[3] Likewise, our axioms and characterization theorem do not presuppose the existence of a prior over the states of nature. Nevertheless, given a prior q over Θ, an experiment induces a distribution over posteriors p, making p a random variable. Under this formulation, the LLR cost (1) of an experiment can be represented as the expected change of the function

$$F(p) = \sum_{i,j\in\Theta} \beta_{ij} \frac{p_i}{q_i} \log\left(\frac{p_i}{p_j}\right)$$

from the prior q to the posterior p induced by the signal. That is, the cost of an experiment equals

$$\mathbb{E}\left[F(p) - F(q)\right],$$

where the expectation is taken with the distribution of posterior beliefs induced by the experiment and the prior. This establishes that LLR cost is posterior separable, and

[3] Blackwell experiments have been studied both within and outside the Bayesian framework. See, for instance, Le Cam (1996) for a review of the literature on Blackwell experiments.

makes it possible to apply techniques and insights derived for posterior separable cost functions (Caplin and Dean 2013; Caplin, Dean, and Leahy 2018).

Relation to Mutual Information Cost.—Following the seminal work of Sims (2003, 2010) on rational inattention, cost functions based on mutual information have been commonly used in applications; Mackowiak, Matějka, and Wiederholt (2018) review the literature on rational inattention. Mutual information costs are defined as the expected change

$$\mathbb{E}\big[H(q) - H(p)\big]$$

of the Shannon entropy $H(p) = -\sum_{i\in\Theta} p_i \log p_i$ between the decision-maker's prior belief q and posterior p. Equivalently, in this formulation, the cost of an experiment is given by the mutual information between the state of nature and the signal.[4] One of the main differences between mutual information and the LLR cost, is that the first is subadditive rather than additive (see, e.g., Lindley 1956), so that the cost of n independent copies of an experiment is a strictly concave function of n. In applications, the LLR cost function leads to predictions which are qualitatively different from those induced by mutual information cost. We illustrate the differences in Section IV and Section V.

Examples and Applications.—In Section V we apply the LLR cost function to information acquisition problems and derive a number of predictions. Our applications include binary prediction problems, where a decision-maker needs to predict whether the state is above or below a given threshold. An example of this is an analyst trying to predict which party will obtain the majority of votes in an election. Another example is a perception task where a subject is asked to observe a number of dots of two different colors on a screen, and must predict which color is predominant.[5]

We show that in binary prediction problems the decision-maker is strictly more likely to make the correct choice when the quantity to be predicted is farther from the desired threshold, under general assumptions on the coefficients (β_{ij}). For example, it is harder for the agent to predict the winner in a close election than in an election where one of the candidates has a large lead. Moreover, we show that under the specification $\beta_{ij} = \kappa/(i-j)^2$, the decision-maker's probability of a choosing an action is a sigmoidal function of the state—a prediction in line with psychometric evidence on perception tasks.

This and other examples illustrate how the LLR cost function leads to optimal choice probabilities that take into account the difficulty of distinguishing between states. While intuitive, this property is ruled out by cost functions such as mutual information that treat states symmetrically.

[4] Related specifications discussed in the literature include models where the decision-maker can acquire, for free, any experiment whose mutual information is below an upper bound (Sims 2003), as well as costs that are increasing transformation of mutual information (Denti 2022).

[5] The two examples have a similar structure but are, of course, quite different in terms of data collection since perception tasks are usually performed with experimental subjects in controlled environments.

Scope and Limitations.—There are many applications where the our additivity assumption is violated, and so the LLR cost function is inadequate. A stark case, which we discuss in the next section, is that of experiments that completely rules out a state; these would have infinite LLR cost. Thus our framework is incompatible with partitional information structures, which are an important modeling tool. Moreover, the fact that our representation has a number of parameters that grows with the number of states makes calculations and identification more difficult.

A natural question is how the LLR cost can be applied in dynamic settings in which agents decide sequentially what information to acquire. As discussed in depth by Bloedel and Zhong (2020), it is impossible—under reasonable assumptions—to have a cost function that satisfies the assumption of constant marginal costs and is independent of the prior of the decision-maker. This is a subtle issue which we explore in more detail in Section VI.

I. Model

A decision-maker acquires information on an unknown state of nature belonging to a finite set Θ. Elements of Θ will be denoted by i, j, k, etc. Following Blackwell (1951), we model the information acquisition process by means of *experiments*. An experiment $\mu = (S, (\mu_i)_{i\in\Theta})$ consists of a set S of signal realizations equipped with a sigma-algebra Σ, and for each state $i \in \Theta$ a probability measure μ_i defined on (S, Σ). The set S represents the possible outcomes of the experiment, and each measure μ_i describes the distribution of outcomes when the true state is i.

We assume throughout that the measures (μ_i) are mutually absolutely continuous, so that each derivative (i.e., ratio between densities) $\frac{\mathrm{d}\mu_i}{\mathrm{d}\mu_j}$ is finite almost everywhere. In the case of finite signal realizations these derivatives are simply equal to ratio between probabilities $\mu_i(s)/\mu_j(s)$.[6]

Given an experiment μ, we denote by

$$\ell_{ij}(s) = \log \frac{\mathrm{d}\mu_i}{\mathrm{d}\mu_j}(s)$$

the log-likelihood ratio between states i and j upon observing the realization s. We define the vector

$$\left(\ell_{ij}(s)\right)_{i,j\in\Theta}$$

of log-likelihood ratios among all pairs of states. The distribution of ℓ depends on the true state generating the data. Given an experiment μ, we denote by $\bar{\mu}_i$ the distribution of ℓ conditional on state i.[7]

We restrict our attention to experiments where the induced log-likelihood ratios (ℓ_{ij}) have finite moments. That is, experiments such that for every state i and every vector of integers $\alpha \in \mathbb{N}^{\Theta}$ the expectation $\int_S |\prod_{k\neq i} \ell_{ik}^{\alpha_k}| \mathrm{d}\mu_i$ is finite. We denote by

[6] This assumption means that no signal can ever rule out any state, and in particular can never completely reveal the true state.

[7] The measure $\bar{\mu}_i$ is defined as $\bar{\mu}_i(A) = \mu_i(\{s : (\ell_{ij}(s)) \in A\})$ for every measurable $A \subseteq \mathbb{R}^{\Theta\times\Theta}$.

$\mathcal{E}$ the class of all such experiments.[8] The restriction to $\mathcal{E}$ is a technical condition that rules out experiments whose log-likelihood ratios have very heavy tails, but, to the best of our knowledge, includes all (not fully revealing) experiments commonly used in applications.

The cost of producing information is described by an *information cost function*

$$C : \mathcal{E} \to \mathbb{R}_+$$

assigning to each experiment $\mu \in \mathcal{E}$ its cost $C(\mu)$. In the next section we introduce and characterize four basic properties for information cost functions.

A. *Axioms*

Our first axiom postulates that the cost of an experiment should depend only on its informational content. For instance, it should not be sensitive to the way signal realizations are labeled. In making this idea formal we follow Blackwell (1951, Section IV).

Let $q \in \mathcal{P}(\Theta)$ be the uniform prior assigning equal probability to each element of Θ.[9] Let μ and ν be two experiments, inducing the distributions over posteriors π_μ and π_ν given the uniform prior q. Then μ dominates ν in the Blackwell order if

$$\int_{\mathcal{P}(\Theta)} f(p) \, \mathrm{d}\pi_\mu(p) \geq \int_{\mathcal{P}(\Theta)} f(p) \, \mathrm{d}\pi_\nu(p)$$

for every convex function $f : \mathcal{P}(\Theta) \to \mathbb{R}$. As is well known, dominance with respect to the Blackwell order is equivalent to the requirement that in any decision problem, a Bayesian decision-maker achieves a (weakly) higher expected utility when basing her action on μ rather than ν. We say that two experiments are *Blackwell equivalent* if they dominate each other.

It is natural to require the cost of information to be increasing in the Blackwell order. For our main result, it is sufficient to require that any two experiments that are Blackwell equivalent lead to the same cost. Nevertheless, it will turn out that our axioms imply the stronger property of Blackwell monotonicity, as shown by Proposition 1 below.

AXIOM 1: *If μ and ν are Blackwell equivalent, then $C(\nu) = C(\mu)$.*

The lower envelope of a cost function assigns to each μ the minimum cost of producing an experiment that is Blackwell equivalent to μ. If experiments are optimally chosen by a decision-maker then we can, without loss of generality, identify a cost function with its lower envelope. This results in a cost function for which Axiom 1 is automatically satisfied.

[8] We refer to $\mathcal{E}$ as a class, rather than a set, since Blackwell experiments do not form a well-defined set. In doing so we follow a standard convention in set theory (see, for instance, Jech 2013, p. 5).

[9] Throughout the paper, $\mathcal{P}(\Theta)$ denotes the set of probability measures on Θ identified with their representation in $\mathbb{R}^\Theta$, so that for every $q \in \mathcal{P}(\Theta)$, q_i is the probability of the state i.

For the next axiom, we study the cost of performing multiple independent experiments. Given two experiments $\mu = (S, (\mu_i))$ and $\nu = (T, (\nu_i))$ we define their product

$$\mu \otimes \nu = \big(S \times T, (\mu_i \times \nu_i)\big),$$

where $\mu_i \times \nu_i$ denotes the product of the two measures.[10] Under the experiment $\mu \otimes \nu$, the realizations of both experiments μ and ν are observed, and the two observations are independent conditional on the state. To illustrate, suppose μ and ν consist of drawing a random sample from two possible populations. Then $\mu \otimes \nu$ is the experiment where two independent samples, one for each population, are collected.

Our second axiom states that the cost function is additive with respect to combining independent experiments:

AXIOM 2: *The cost of performing two independent experiments is the sum of their costs:*

$$C(\mu \otimes \nu) = C(\mu) + C(\nu) \quad \textit{for all } \mu \textit{ and } \nu.$$

An immediate implication of Axioms 1 and 2 is that a completely uninformative experiment has zero cost. This follows from the fact that an uninformative experiment μ is Blackwell equivalent to the product experiment $\mu \otimes \mu$.

In many settings, an experiment can sometimes fail to produce new evidence. The next axiom states that the cost of an experiment is linear in the probability that it will generate information. Given μ, we define a new experiment, which we call a *dilution* of μ and denote by $\alpha \cdot \mu$. In this new experiment, with probability α the experiment μ is produced, and with probability $1 - \alpha$ a completely uninformative signal is observed. Formally, given $\mu = (S, \mu_i))$, fix a new signal realization $o \notin S$ and a probability $\alpha \in [0,1]$. We define

$$\alpha \cdot \mu = \big(S \cup \{o\}, (\nu_i)\big),$$

where $\nu_i(E) = \alpha \mu_i(E)$ for every measurable $E \subseteq S$, and $\nu_i(\{o\}) = 1 - \alpha$. The next axiom specifies the cost of such an experiment:

AXIOM 3: *The cost of a dilution $\alpha \cdot \mu$ is linear in the probability α:*

$$C(\alpha \cdot \mu) = \alpha C(\mu) \quad \textit{for every } \mu \textit{ and } \alpha \in [0,1].$$

Our final assumption is a continuity condition. We first introduce a (pseudo-) metric over $\mathcal{E}$. Recall that for every experiment μ, $\bar{\mu}_i$ denotes its distribution of log-likelihood ratios conditional on state i. We denote by d_{tv} the total-variation distance.[11] Given a vector $\alpha \in \mathbb{N}^{\Theta}$, let $M_i^{\mu}(\alpha) = \int_S |\prod_{k \neq i} \ell_{ik}^{\alpha_k}| d\mu_i$ be the α-moment

[10] When the set of signal realizations is finite, the measure $\mu_i \times \nu_i$ assigns to each realization (s,t) the probability $\mu_i(s)\,\nu_i(t)$.

[11] That is, $d_{tv}(\bar{\mu}_i, \bar{\nu}_i) = sup|\bar{\mu}_i(A) - \bar{\nu}_i(A)|$, where the supremum is over all measurable subsets of $\mathbb{R}^{\Theta \times \Theta}$.

of the vector of log-likelihood ratios $(\ell_{ik})_{k\neq i}$. Given an upper bound $N \geq 1$, we define the distance:

$$d_N(\mu,\nu) = \max_{i\in\Theta} d_{tv}(\bar{\mu}_i,\bar{\nu}_i) + \max_{i\in\Theta} \max_{\alpha\in\{0,\ldots,N\}^n} \left|M_i^{\mu}(\alpha) - M_i^{\nu}(\alpha)\right|.$$

According to the metric dN, two experiments μ and ν are close if, for each state i, the induced distributions of log-likelihood ratios are close in total variation and, in addition, have similar moments, for any moment α lower or equal to $(N,\ldots,N)$.

AXIOM 4: *For some $N \geq 1$ the function C is uniformly continuous with respect to d_N.*

As is well known, convergence with respect to the total-variation distance is a demanding requirement, as compared to other topologies such as the weak topology. So, continuity with respect to d_{tv} is a relatively mild assumption. Continuity with respect to the stronger metric d_N is, therefore, an even milder assumption.[12] As we show in Theorem 6 in the online Appendix, our characterization holds for the case of two states and bounded experiments even if one only imposes Blackwell monotonicity, Axiom 2 and Axiom 3, without requiring continuity.

B. *Discussion*

We now discuss the interpretation of our axioms as well as some limitations imposed by our modeling assumptions. Axiom 2 has a simple interpretation. Consider the classical problem of learning the bias of a coin by flipping it multiple times. This experiment could correspond to the act of surveying customers, who either like a product or not, in order to learn whether the product is popular. It could also represent a political party surveying voters to discover the appeal of a potential candidate.

Suppose the coin either yields heads 80 percent of the time or tails 80 percent of the time and that either bias is equally likely. We compare the cost of observing a single coin flip versus a long sequence of coin flips. Under the additivity axiom, the cost of observing k coin flips is linear in k.

Additivity assumptions in the spirit of Axiom 2 have appeared in multiple parametric models of information acquisition. A standard assumption in Wald's classic model of sequential sampling and its variations is that the cost of acquiring n independent samples is linear in n(see, e.g., Wald 1945; Arrow, Blackwell, and Girshick 1949). A similar condition appears in the continuous-time formulation of the sequential sampling problem, where the information structure consists of observing a signal with Brownian noise over a time period of length t, under a cost that is linear in t (Dvoretzky, Kiefer, and Wolfowitz 1953; Chan et al. 2017; Morris and Strack 2018). Likewise, in static models where information is acquired by means of normally

[12] We discuss this topology in detail in the online Appendix. Any information cost function that is continuous with respect to the metric d_N satisfies Axiom 1. For expositional clarity, we maintain the two axioms as separate throughout the paper.

distributed experiments, a standard specification is that the cost of an experiment is inversely proportional to its variance (see, e.g., Wilson 1975). This amounts to an additivity assumption, since the product of two independent normal experiments is Blackwell equivalent to a normal experiment whose precision is the sum of the original precisions. Underlying these different models is the notion that the cost of an additional independent experiment is constant. Axiom 2 captures this idea in a nonparametric context, with no a priori restrictions over the domain of feasible experiments.

Axiom 3 expresses the idea that the marginal cost of increasing the probability of success of an experiment is constant. The axiom is implied by posterior separability—the standard assumption in the literature for cost functions over experiments.[13] It is however, a strictly weaker assumption. We also note that for proving our results it suffices to restrict this axiom to $\alpha = 1/2$.[14]

The domain of our cost function rules out experiments that with positive probability allow the decision-maker to be certain that a state did not happen. Such experiments, if included in the domain, would have infinite cost under our axioms.[15] While this is not special to our framework—the same issue applies to Wald's model and others—it is nevertheless an important limitation, since information structures that rule out states with certainty are a common modeling tool. An example are partitional information structures, which are standard in information economics. A disadvantage of the LLR cost function is that it cannot be applied in such settings.

To gain some intuition for the sort of experiments that are ruled out, consider an urn containing 100 balls. Suppose there are only two states: either all balls are red, or all balls are blue. In this case, sampling from the urn perfectly reveals the state, and thus such an experiment cannot be accommodated by the LLR cost. Indeed, it conflicts with the constant marginal cost assumption: if the experiment had finite cost, then repeating it twice would have twice the cost. But repeating the experiment

[13] A posterior separable cost function is affine with respect to the distribution of beliefs induced by an experiment. The distribution of beliefs induced by the diluted experiment $\alpha \cdot \mu$ is a convex combination that puts weight α on the distribution generated by μ and weight $1 - \alpha$ on the prior. Thus, under posterior separability the cost of $\alpha \cdot \mu$ is affine in α.

[14] This axiom admits an additional interpretation. Suppose the decision-maker is allowed to randomize her choice of experiment. Then, the property

$$C(\alpha \cdot \mu) \leq \alpha C(\mu) \tag{2}$$

ensures that the cost of the diluted experiment $\alpha \cdot \mu$ is not greater than the expected cost of performing μ with probability α and collecting no information with probability $1 - \alpha$. Hence, if (2) were violated, the experiment $\alpha \cdot \mu$ could be replicated at a strictly lower cost through a simple randomization by the decision-maker. Now assume Axiom 2 holds, and the decision-maker is allowed to perform independent copies of the diluted experiment $\alpha \cdot \mu$ until it succeeds. Then, the converse inequality

$$C(\alpha \cdot \mu) \geq \alpha C(\mu)$$

ensures that the cost $C(\mu)$ of an experiment is not greater than the expected cost $(1/\alpha)C(\alpha \cdot \mu)$ of performing the experiment $\alpha \cdot \mu$ until it succeeds.

[15] For example, if a cost function C is Blackwell monotone, additive, and assigns strictly positive cost to at least one experiment μ that is not perfectly revealing, then it must assign infinite cost to a perfectly revealing experiment. Indeed, by Blackwell monotonicity, the cost of the n-times repeated experiment $\mu^{\otimes n}$ must always be below the cost of a perfectly informative experiment. By additivity, $C(\mu^{\otimes n}) = nC(\mu)$, and thus a perfectly informative experiment must have infinite cost.

does not provide any additional information, since one sample is enough to reveal the state. Thus, the constant marginal cost assumption fails in this example.

Suppose instead the urn contains either 1 blue ball and 99 red balls, or 1 red ball and 99 blue ones. In this case, drawing from the urn is an experiment that does not exclude states with certainty, and fits with the assumption of additivity. As the number of samples grows, the decision-maker obtains more and more accurate statistical evidence of the true state, but without ever reaching full certainty.

II. Representation

THEOREM 1: *An information cost function C satisfies Axioms 1–4 if and only if there exists a collection* $(\beta_{ij})_{i,j\in\Theta,i\neq j}$ *in* $\mathbb{R}_+$ *such that for every experiment* $\mu = (S,(\mu_i))$,

$$C(\mu) = \sum_{i,j} \beta_{ij} \int_S \log \frac{\mathrm{d}\mu_i}{\mathrm{d}\mu_j}(s)\mathrm{d}\mu_i(s). \tag{3}$$

Moreover, the collection (β_{ij}) *is unique given C.*

We refer to a cost function that satisfies Axioms 1–4 as an *LLR* cost. As shown by the theorem, this class of information cost functions is uniquely determined up to the parameters (β_{ij}). The expression $\int_S \log(\mathrm{d}\mu_i/\mathrm{d}\mu_j)\mathrm{d}\mu_i$ is the Kullback-Leibler divergence $D_{\mathrm{KL}}(\mu_i \| \mu_j)$ between the two distributions, a well understood and tractable measure of informational content (Kullback and Leibler 1951). The representation (3) can be rewritten as

$$C(\mu) = \sum_{i,j} \beta_{ij} D_{\mathrm{KL}}(\mu_i \| \mu_j).$$

A higher value of $D_{\mathrm{KL}}(\mu_i \| \mu_j)$ describes an experiment which, conditional on state i, produces stronger evidence in favor of state i compared to j, as represented by a higher expected value of the log-likelihood ratio $\log \mathrm{d}\mu_i/\mathrm{d}\mu_j$. The coefficient β_{ij} thus measures the marginal cost of increasing the expected log-likelihood ratio between states i and j, conditional on i, while keeping all other expected log-likelihood ratios fixed.[16]

The specification of the parameters (β_{ij}) must of course depend on the particular application at hand. Consider, for instance, a doctor who must choose a treatment for a patient displaying a set of symptoms, and who faces uncertainty regarding their cause. In this example, a state of nature i represents a possible pathology affecting the patient. In order to distinguish between two possible diseases i and j it is necessary to collect samples and run tests, whose costs will depend on factors that are specific to the two conditions, such as their similarity, or the prominence of their physical manifestations. These differences in costs can then be reflected by the coefficients β_{ij} and β_{ji}. For example, suppose that i and i' are two types of viral

[16] As we formally show in Lemma 2 in the online Appendix, this operation of increasing a single expected log-likelihood ratio while keeping all other expectations fixed is well defined: for every experiment μ and every $\varepsilon > 0$, if $D_{\mathrm{KL}}(\mu_i \| \mu_j) > 0$ then there exists a new experiment ν such that $D_{\mathrm{KL}}(\nu_i \| \nu_j) = D_{\mathrm{KL}}(\mu_i \| \mu_j) + \varepsilon$, and all other divergences are equal. Hence the difference in cost between ν and the experiment μ is given by β_{ij} times the difference ε in the expected log-likelihood ratio. The result formally justifies the interpretation of each coefficient β_{ij} as a marginal cost.

infections, k is a bacterial infection, and i and i' are difficult to tell apart, but telling i and k apart is easier. This can be captured by setting $\beta_{ii'} > \beta_{ik}$. In Section VII we discuss environments where the coefficients might naturally be assumed to be asymmetric, in the sense that $\beta_{ij} \neq \beta_{ji}$.[17] In environments where no pair of states is a priori harder to distinguish than another, a simple choice is to set all the coefficients (β_{ij}) to be equal.[18] Finally, in the next section we propose a specific functional form in the more structured case where states represent a one-dimensional quantity.

We end this section by noting that the LLR cost function is monotone with respect to the Blackwell order:

PROPOSITION 1: *Let μ and ν be experiments such that μ Blackwell dominates ν. Then every LLR cost C satisfies $C(\mu) \geq C(\nu)$.*

III. Learning about a One-Dimensional State

Many information acquisition problems involve learning about a one-dimensional characteristic, so that each state i is a real number. In macroeconomic applications, the state may represent the inflation rate. In perceptual experiments, the state can correspond to the number of red/blue dots on a screen. In a polling problem, the state may correspond to the number of voters voting for a given party. Alternatively, i might represent a physical quantity to be measured.

In this section we propose a choice of parameters (β_{ij}) for one-dimensional information acquisition problems. Given a problem where each state $i \in \Theta \subset \mathbb{R}$ is a real number, we propose to set each coefficient β_{ij} to be equal to $\kappa/(i-j)^2$ for some constant $\kappa \geq 0$. Each β_{ij} is therefore inversely proportional to the squared distance between the corresponding states i and j. Under this specification, two states that are closer to each other are harder to distinguish.

The main result of this section shows that this choice of parameters captures two main hypotheses: (i) the difficulty of producing a signal that allows to distinguish between states i and j is a function only of the distance $|i-j|$ between the two, and (ii) the cost of a noisy measurement of the state with standard normal error is the same across information acquisition problems. Both assumptions take as a working hypothesis that the cost of making a measurement depends only on its precision, and not on the other aspects of the model, such as the set of states Θ. For example, the cost of measuring the height of a person with a given instrument does not depend on whether the person's height is known to be in $\Theta = \{190, \ldots, 210\}$ or $\Theta' = \{160, \ldots, 180\}$.

Let W be a nonempty open interval of $\mathbb{R}$; we think of this set as the range of reasonable values of the state, where our hypotheses apply. We denote by $\mathcal{T}$ the collection of finite subsets of W with at least two elements. Each set $\Theta \in \mathcal{T}$ represents the set of states of nature in a different, one-dimensional, information acquisition problem. To simplify the language, we refer to each Θ as a *problem*. For each $\Theta \in \mathcal{T}$

[17] Since we do not impose symmetry axioms, it is in a sense a natural finding that the LLR cost function can capture differences in the costs of learning about different states. It is perhaps more surprising that the cost function has a relatively small set of $n(n-1)$ parameters, where n is the number of states.

[18] An example common in the literature (e.g., Christie 1934) is that of a detective who has to discover which member of a finite group of people committed a violent crime in some isolated setting, such as a train.

we are given an LLR cost function C^{Θ} with coefficients (β_{ij}^{Θ}). The next two axioms formalize the two hypotheses described above by imposing restrictions, across problems, on the cost of information.

The first axiom states that β_{ij}^{Θ}, the marginal cost of increasing the expected log-likelihood ratio between two states i and j is a function of the distance between the two, and is unaffected by changing the values of the other states.

AXIOM A: *For all* $\Theta, \Xi \in \mathcal{T}$ *such that* $|\Theta| = |\Xi|$, *and for all* $i, j \in \Theta$ *and* $k, l \in \Xi$,

$$\text{if} \quad |i - j| = |k - l| \quad \text{then} \quad \beta_{ij}^{\Theta} = \beta_{kl}^{\Xi}.$$

For each $i \in W$ we denote by ζ_i a normal probability measure on the real line with mean i and variance one. Given a problem Θ, we denote by ζ^{Θ} the experiment $(\mathbb{R}, (\zeta_i)_{i \in \Theta})$. This is the canonical experiment consisting of a noisy measurement of the state plus standard normal error. Expressed differently, if $i \in \Theta$ is the true state, then the outcome of the experiment ζ^{Θ} is distributed as $s = i + \varepsilon$, where ε is normally distributed with mean zero and variance one independent of the state. The next axiom states that the cost of such a measurement does not depend on the particular values that the state can take.

AXIOM B: *For all* $\Theta, \Xi \in \mathcal{T}$, $C^{\Theta}(\zeta^{\Theta}) = C^{\Xi}(\zeta^{\Xi})$.

Axioms A and B lead to a simple parametrization for the coefficients of the LLR cost in one-dimensional information acquisition problems:

PROPOSITION 2: *The collection* $C^{\Theta}, \Theta \in \mathcal{T}$, *satisfies Axioms A and B if and only if there exists a constant* $\kappa > 0$ *such that for all* $i, j \in \Theta$ *and* $\Theta \in \mathcal{T}$,

$$\beta_{ij}^{\Theta} = \frac{\kappa}{n(n-1)} \frac{1}{(i-j)^2},$$

where n *is the cardinality of* Θ.

Thus, under Axioms A and B each coefficient β_{ij}^{Θ} is decreasing in the distance between the states, so that distinguishing states that are closer to each other is more costly. Each coefficient is also divided by a factor $n(n-1)$ that normalizes the cost with respect to the number of states. This is an implication of Axiom B, which states that the cost of performing a noisy measurement does not depend on the particular values the state can take. As we show in the proof, the quadratic term $(i-j)^2$ in the expression of the coefficients is related to the assumption, in the same axiom, of normally distributed noise. In the online Appendix we show how the results can be extended to different families of distributions.

Applied to normal experiments, Proposition 2 implies that for any $\Theta \in \mathcal{T}$, a normal experiment with mean i and variance σ^2 has cost $\kappa\sigma^{-2}$ proportional to its precision. Thus, this functional form generalizes a specification often found in the literature—where the cost of a normal experiment is assumed to be proportional to its precision (Wilson 1975)—to arbitrary information structures that are not necessarily normal.

As we will see in Section V, the specification of Proposition 2 allows to compute numerical solutions, and thus can be useful for deriving quantitative predictions. At the same time, this functional form may be too simple to capture certain intuitive comparative statics with respect to changes of the state space. For example, the precision of a measurement made using a measuring tape is quite different when measuring a person's height than when measuring the length of a field. More generally, any measurement instrument has a range of reliability, and as one moves toward the extremes it becomes noisier. We partially address this issue by allowing the state to only take value in some interval $W \subseteq \mathbb{R}$.

Axiom A assumes that only the distance between states determines the cost of an experiment. But in many situations states with a given distance are harder to distinguish at larger scales. Consider, for instance, a subject in a laboratory experiment who is asked to guess the number of pennies in a jar. A problem where this state can take the values either 1 or 2 is easier than a problem where the state can take the values 101 or 102.

Such examples form a well known empirical regularity in psychophysics, known as *Weber's law*, according to which the change in stimulus intensity that is necessary for subjects to exhibit a certain response is a constant fraction of the starting intensity of the stimulus. A way to model such situations is to change the units in which states are measured by applying a logarithmic transformation to the states. This is equivalent to changing Axiom A to consider ratios between states instead of differences, and changing Axiom B to consider log-normal measurement errors instead of normal. The resulting coefficients are

$$\beta_{ij}^{\Theta} = \frac{\kappa}{n(n-1)} \frac{1}{(\log i/j)^2}.$$

This results in predictions in line with Weber's law, making it easier to distinguish 1 from 2 than 101 from 102.

IV. Illustrative Examples

A. *LLR Cost for Normal and Binary Experiments*

Closed form solutions for the Kullback-Leibler divergence between standard distributions such as normal, exponential or binomial, are readily available. This makes it immediate to compute the cost of common parametric families of experiments.

Normal Experiments.—Consider a normal experiment $\mu^{m,\sigma}$ according to which the signal s is given by

$$s = m_i + \varepsilon,$$

where the mean $m_i \in \mathbb{R}$ depends on the true state i, and ε is state independent and normally distributed with standard deviation σ. In this example, each m_i is a feature of the information structure: choosing an experiment where the distances between states $|m_i - m_j|$ are higher provides stronger information about the states.

By substituting (3) with the well-known expression for the Kullback-Leibler divergence between normal distributions, we obtain that the cost of such an experiment is given by

$$C(\mu^{m,\sigma}) = \sum_{i,j} \beta_{ij} \frac{(m_j - m_i)^2}{2\sigma^2}. \tag{4}$$

The cost is decreasing in the variance σ^2, as one may expect. Increasing β_{ij} increases the cost of an experiment $\mu^{m,\sigma}$ by a factor that is proportional to the squared distance between the means of the two experiments.

Binary Experiments.—Another canonical example is the binary-binary setting in which the set of states is $\Theta = \{H,L\}$, and the experiment $\nu^p = (S,(\nu_i))$ is also binary: $S = \{0,1\}$, $\nu_H = B(p)$ and $\nu_L = B(1-p)$ for some $p > 1/2$, where $B(p)$ is the Bernoulli distribution on $\{0,1\}$ assigning probability p to 1. In this case, the cost increases in p and given by

$$C(\nu^p) = (\beta_{HL} + \beta_{LH})\left[p \log \frac{p}{1-p} + (1-p)\log \frac{1-p}{p}\right]. \tag{5}$$

B. *Hypothesis Testing*

In this section, we apply the LLR cost to a standard hypothesis testing problem. We study a decision-maker performing an experiment with the goal of learning about a hypothesis, i.e., whether the state is in a subset $H \subset \Theta$.

We consider an experiment that reveals with some probability whether the hypothesis is true or not, and study how its cost depends on the structure of H. For a given hypothesis H and a precision α let μ be the symmetric binary experiment with signal realizations $S = \{H, H^c\}$, where H^c denotes the complement of H:

$$\mu_i(s) = \begin{cases} \alpha & \text{for} \quad i \in s \\ 1-\alpha & \text{for} \quad i \notin s \end{cases}. \tag{6}$$

Conditional on any state, this experiment yields a correct signal with probability α. Under LLR cost, the cost of such an experiment is given by

$$\left(\sum_{i \in H, j \in H^c} \beta_{ij} + \beta_{ji}\right)\left(\alpha \log \frac{\alpha}{1-\alpha} + (1-\alpha)\log \frac{1-\alpha}{\alpha}\right). \tag{7}$$

The first term captures the difficulty of discerning between H and H^c. The harder the states in H and H^c are to distinguish, the larger the sum of the coefficients β_{ij} and β_{ji} will be, and the more costly it will thus be to learn whether the hypothesis H is true. The second term is monotone in the precision α and is independent of the hypothesis. We illustrate with an example how this captures the fact that testing two different hypotheses can lead to very different costs even if they involve the same number of states.

Learning about the GDP.—For concreteness, we take a state to be a natural number i in the interval $\Theta = \{20{,}000, \ldots, 80{,}000\}$, representing the current US GDP per capita. We consider the following two hypotheses:

(H1) The GDP is above 50,000.

(H2) The GDP is an even number.

Intuitively, producing enough information to answer with high accuracy whether H1 is true should be less expensive than producing enough information to answer whether H2 is true, a practically impossible task. Our model captures this intuition: As the state is one-dimensional, we set $\beta_{ij} = \kappa/(i-j)^2$ following Section III; the same qualitative conclusion will hold as long as β_{ij} is strictly decreasing in the distance $|i-j|$. Then

$$\sum_{i \in H1, j \in H1^c} \beta_{ij} + \beta_{ji} \approx 22\kappa, \quad \sum_{i \in H2, j \in H2^c} \beta_{ij} + \beta_{ji} \approx 148033\kappa.$$

That is, learning whether the GDP is even or odd is by several orders of magnitude more costly than learning whether the GDP is above or below 50,000.[19]

It is useful to compare these observations with the results that would be obtained under mutual information and a uniform prior on Θ. In such a model, the cost of a symmetric binary experiment with precision α is determined solely by the cardinality of H. In particular, under mutual information learning whether the GDP is above or below 50,000 is equally costly as learning whether it is even or odd. This is a well-known property of cost functions that are invariant with respect to a relabeling of the states.

V. Information Acquisition in Decision Problems

In this section we study the implications of the LLR cost function for decision problems. We consider a decision-maker choosing an action a from a finite set A. The payoff from a depends on the state i and is given by $u(a,i)$. The agent is endowed with a full-support prior q over the set of states. Before making her choice, the agent can acquire an experiment $\mu \in \mathcal{E}$ at cost $C(\mu)$, where C is an LLR cost function where the coefficients (β_{ij}) are assumed to be positive.

As is well known, for a cost function that is monotone with respect to the Blackwell order, it is without loss of generality to restrict attention to experiments where the set of realizations S equals the set of actions A, and to assume that upon observing a signal $s = a$ the decision-maker will choose the action recommended by the signal. Throughout this section, we will therefore identify an experiment μ with a vector of probability measures over actions $\mu \in \mathcal{P}(A)^n$ where $n = |\Theta|$.

[19] Beyond the challenge of learning about the state, which is the focus of this paper, it might be computationally difficult to determine the set that corresponds to a given hypothesis. Consider, for example, the hypothesis (H1) The number of pages in this manuscript is an even number, versus the hypothesis (H2) The number of pages in this manuscript is greater than $\sqrt{4000}$. The relative "distance" properties of the states are in both cases exactly the same as in the GDP example, but the cost of telling states apart is considerably higher in the high-distance case than in the low distance one. We thank the editor for suggesting this example.

An optimal experiment $\mu^\star = (\mu_i^\star)$ solves

$$\mu^\star \in \operatorname*{arg\,max}_{\mu \in \mathcal{P}(A)^n} \sum_{i \in \Theta} q_i \Big(\sum_{a \in A} \mu_i(a) u(a,i) \Big) - C(\mu). \tag{8}$$

Hence, the optimal action a is chosen in state i with probability $\mu_i^\star(a)$. The maximization problem (8) is well behaved: the maximand is upper-semicontinuous and concave (see Proposition 10 in the online Appendix), and there always exists an optimal solution.[20] Thus, an optimal experiment can be found by applying standard methods in concave optimization.

It is without loss of generality to restrict attention to choice probabilities where an action that is chosen with strictly positive probability in one state is chosen with strictly positive probability in every state, since otherwise the experiment is not in the domain $\mathcal{E}$.

A. *Implications for Optimal Choice Probabilities*

We obtain a characterization of the decision-maker's optimal choice probabilities. The characterization is based on the study of first-order conditions, and is therefore analogous to that obtained by Matějka and McKay (2015) for the case of mutual information cost.

The result is based on a standard economic intuition. For choice probabilities to be optimal, the marginal benefit of choosing an action a marginally more often than a different action b must exactly offset its marginal cost. Formally, given a vector μ of choice probabilities, we denote by $\mathrm{supp}(\mu)$ the support of μ, i.e., the set of actions that are played with strictly positive probability under μ.[21] Given two actions a and b in the support of μ, consider perturbing μ by increasing $\mu_i(a)$ while decreasing $\mu_i(b)$ by the same amount. The marginal benefit of this perturbation is denoted by $\mathrm{MB}_i(a,b)$ and is equal to

$$\mathrm{MB}_i(a,b) = q_i\big[u(a,i) - u(b,i)\big].$$

Such a transfer of probabilities has an effect on the information cost of the experiment μ. This is given by the expression

$$\mathrm{MC}_i(a,b) = \sum_{j \in \Theta} \beta_{ij} \left(\log \frac{\mu_i(a)}{\mu_j(a)} - \log \frac{\mu_i(b)}{\mu_j(b)} \right) - \sum_{j \in \Theta} \beta_{ji} \left(\frac{\mu_j(a)}{\mu_i(a)} - \frac{\mu_j(b)}{\mu_i(b)} \right). \tag{9}$$

It measures the change in information acquisition cost necessary to choose action a marginally more often and action b marginally less often. For the choice probabilities μ to be chosen optimally, this change in information cost must equal the difference $q_i[u(i,a) - u(i,b)]$ in expected benefits. This is the content of the next proposition.

[20] To establish existence of an optimal solution, recall that the Kullback-Leibler divergence $D_{\mathrm{KL}}:\mathcal{P}(A) \times \mathcal{P}(A) \to [0,\infty]$ is a lower-semicontinuous function (Dupuis and Ellis 2011, Lemma 1.4.3). The maximand in (8), being a sum of upper-semicontinuous functions, is upper-semicontinuous. Since $\mathcal{P}(A)^n$ is compact, the problem admits a solution.

[21] That is, $\mathrm{supp}(\mu) = \{a \in A : \mu_i(a) > 0 \text{ for some } i \in \Theta\}$.

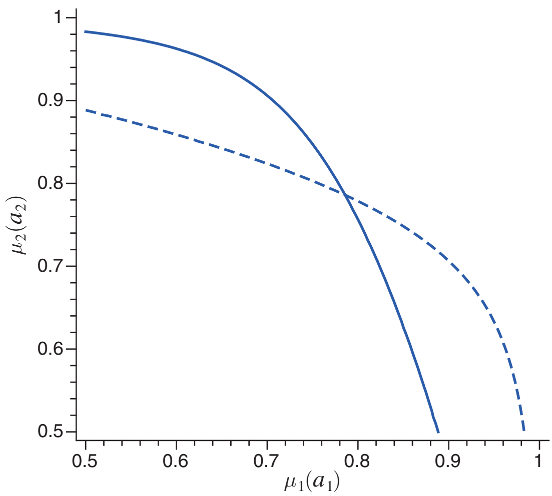

FIGURE 1

Notes: A decision problem where $\Theta = \{1,2\}$ and $A = \{a_1, a_2\}$, the prior q is uniform, $\beta_{12} = \beta_{21} = 1$, and payoffs are $u_1(a_1) = u_2(a_2) = 3$ and $u_1(a_2) = u_2(a_1) = 0$. The solid line is the locus of choice probabilities such that $\mathrm{MB}_1(a_1,a_2) = \mathrm{MC}_1(a_1,a_2)$. The dotted line is the locus where $\mathrm{MB}_2(a_2,a_1) = \mathrm{MC}_2(a_2,a_1)$. The optimal vector of choice probabilities is given by the intersection of the two curves.

PROPOSITION 3: *Let* $\mu = (\mu_i)_{i\in\Theta}$ *be the vector of choice probabilities that solves the optimization problem* (8). *Then, for every state* $i \in \Theta$ *it holds that*

$$\mathrm{MB}_i(a,b) = \mathrm{MC}_i(a,b) \quad \textit{for all} \quad a,b \in \mathrm{supp}(\mu). \tag{10}$$

Figure 1 illustrates this result in a simple decision example with two states and two actions where the decision-maker's goal is to match the state. Proposition 3 characterizes the optimal choice probabilities in terms of necessary first-order conditions. These conditions are in general not sufficient, because they do not verify that the support of μ is optimal. In the case of mutual information, Caplin, Dean, and Leahy (2016) and Denti, Marinacci, and Montrucchio (2020) give a characterization of the set of actions that are taken with positive probability, and arrive at first-order conditions that are both sufficient and necessary. We do not know whether analogous first-order conditions can be obtained for the LLR cost function.

B. *Continuity of Choice Probabilities*

A feature of the LLR cost is its ability to model the fact that closer states are harder to distinguish, in the sense that acquiring information that finely discriminates between them is more costly. This, in turn, suggests that choice probabilities cannot vary abruptly across nearby states.

To formalize this intuition, we assume that the state space Θ is endowed with a distance $d:\Theta \times \Theta \to \mathbb{R}$. We say that *nearby states are hard to distinguish* if for all $i,j \in \Theta$

$$\beta_{ij} \geq \frac{1}{d(i,j)^2}. \tag{11}$$

Under this assumption the cost of acquiring information that discriminates between states i and j is high for states that are close to each other. Our next result shows that when nearby states are hard to distinguish, the optimal choice probabilities are Lipschitz continuous in the state: the agent will choose actions with similar probabilities in similar states. For this result, we denote by $\|u\| = \max_{i,a}|u(a,i)|$ the norm of the decision-maker's utility function.

PROPOSITION 4 (Continuity of Choice): *Suppose that nearby states are hard to distinguish. Then the optimal choice probabilities $\mu^\star$ solving (8) are uniformly Lipschitz continuous with constant $\sqrt{\|u\|}$, i.e., satisfy*

$$\sum_{a\in A}\left|\mu_i^\star(a) - \mu_j^\star(a)\right| \leq \sqrt{\|u\|}\, d(i,j) \quad \textit{for all} \quad i,j \in \Theta. \tag{12}$$

Lipschitz continuity is a standard notion of continuity in discrete settings, such as the one of this paper, where the relevant variable i takes finitely many values. A crucial feature of the bound (12) is that the Lipschitz constant depends only on the norm $\|u\|$ of the utility function, independently of the exact form of the coefficients (β_{ij}), and of the number of states.[22] In addition, assumption (11) can be generalized to arbitrary ordinal transformations of the distance d. The proof of Proposition 4 shows that if the coefficients satisfy $\beta_{ij} \geq 1/f(d(i,j))^2$ for a monotone increasing function f, then the conclusion of the proposition holds with the right-hand side of (12) replaced with $\sqrt{\|u\|}\, f(d(i,j))$.

This result highlights a contrast between the predictions of mutual information cost and LLR cost. Mutual information predicts behavior that displays a discontinuity with respect to the state (see Section VE for an example). Under LLR cost, when nearby states are harder to distinguish, the change in choice probabilities across states can be bounded by the distance between them.

This difference has stark implications in coordination games. Morris and Yang (2016) study information acquisition in coordination problems. In their model, continuity of the choice probabilities with respect to the state leads to a unique equilibrium; if continuity fails, then there are multiple equilibria. This suggests that different choices of information cost can lead to different predictions in coordination games and their economic applications.

C. *Comparative Statics with Respect to the Coefficients β_{ij}*

While so far we have focused on the effect that the coefficients β_{ij} have on the cost of a given experiment, we now address the question of their effect on behavior. The next proposition is a comparative statics result describing how choice probabilities vary with the parameters β_{ij}.

[22] Proposition 4 suggests that the analysis of choice probabilities might be extended to the case where the set of states Θ is an interval in $\mathbb{R}$, or, more generally, a metric space. Given a (possibly infinite) state space Θ endowed with a metric, and a sequence of finite discretizations (Θ_n) converging to Θ, the bound (12) implies that if the corresponding sequence of choice probabilities converges, then it must converge to a collection of choice probabilities that are continuous, and moreover Lipschitz.

PROPOSITION 5: *Consider a decision problem, and let* μ *and* μ' *be the optimal choice probabilities obtained under an LLR cost function with coefficients* (β_{ij}) *and* (β'_{ij}), *respectively. Then*

$$\sum_{i \neq j} (\beta'_{ij} - \beta_{ij}) \big(D_{\mathrm{KL}}(\mu'_i \,\|\, \mu'_j) - D_{\mathrm{KL}}(\mu_i \,\|\, \mu_j) \big) \leq 0.$$

All other things equal, an increase of the coefficient β_{ij} decreases the Kullback-Leibler divergence $D(\mu_i \,\|\, \mu_j)$ between the corresponding optimal choice probabilities, and thus makes the decision-maker's behavior more similar in the two states.

Proposition 5 follows from the same logic underlying the law of supply in standard microeconomic models of production. Under the LLR cost function, the decision-maker solves an optimization problem that is mathematically equivalent to a profit maximization problem. Each expected log-likelihood ratio $D_{\mathrm{KL}}(\mu_i \,\|\, \mu_j)$ is an intermediate "input" which accrues to the decision-maker's expected payoff. Each such input is "priced" according to a linear price β_{ij}. The comparative statics described by the result follows from such a linearity property, together with a standard revealed-preference argument.

D. *Identifying the Cost from Observed Choices*

Proposition 3 can be applied to the problem of identifying and testing the LLR model from observed choices. We illustrate this in the context of a simple example. We consider a binary choice problem where we are given two a priori equally likely states $\Theta = \{1, 2\}$. The agent can take one of two actions, a_1 and a_2, and receives a payoff $v > 0$ if the action matches the state and 0 otherwise.

An analyst observes the agent's choice probabilities $(\mu_i(a))_{i \in \Theta, a \in A}$, and is interested in testing if such probabilities are consistent with LLR cost. This is true if there exist coefficients (β_{12}, β_{21}) that satisfy equation (10). The equation simplifies to

$$\text{(13)} \qquad \frac{v}{2} = -\big[\beta_{12}(\log l_1 - \log l_2) + \beta_{21}(l_1 - l_2)\big],$$

$$-\frac{v}{2} = -\big[\beta_{21}(-\log l_1 + \log l_2) + \beta_{12}(1/l_1 - 1/l_2)\big];$$

where $l_1 = \mu_2(1)/\mu_1(1)$ and $l_2 = \mu_2(2)/\mu_1(2)$. Rearranging the conditions above yields that one can infer the information cost parameters (β_{ij}) from her choice probabilities μ as

$$\text{(14)} \qquad \beta_{12} = \frac{v}{2} \frac{l_2 - l_1 + \log \frac{l_1}{l_2}}{\frac{(l_1 - l_2)^2}{l_1 l_2} - \left(\log \frac{l_1}{l_2}\right)^2}, \quad \beta_{21} = \frac{v}{2} \frac{\frac{l_2 - l_1}{l_1 l_2} + \log \frac{l_1}{l_2}}{\frac{(l_1 - l_2)^2}{l_1 l_2} - \left(\log \frac{l_1}{l_2}\right)^2}.$$

For example, if the agent takes the correct action 80 percent of the time in state 1 and 60 percent of the time in state 2, we have that $(\mu_1(1), \mu_1(2), \mu_2(1), \mu_2(2)) = (0.8, 0.2, 0.4, 0.6)$ and the formula above yields that $(\beta_{12}, \beta_{21}) \approx (0.37v, -0.07v)$. As the implied β_{21} is negative these choice probabilities are inconsistent with any LLR cost function and this type of choice behavior would reject our model. In contrast, if the agent takes the correct action 80 percent of the time in state 1

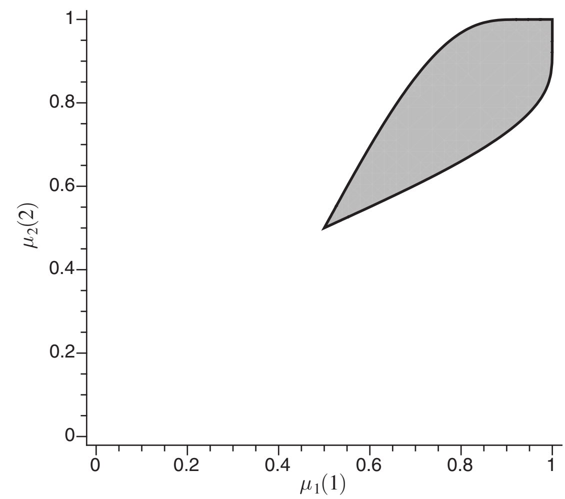

FIGURE 2

Note: The probabilities of choosing correctly in state 1 and state 2 that are consistent with LLR cost.

and 70 percent of the time in state 2, we have that $(\mu_1(1), \mu_1(2), \mu_2(1), \mu_2(2)) = (0.8, 0.2, 0.3, 0.7)$ which implies that $(\beta_{12}, \beta_{21}) \approx (0.18v, 0.03v)$, and thus that this choice behavior can be explained by an LLR cost. Figure 2 more generally depicts all probabilities of choosing correctly in state 1 and state 2 that are consistent with LLR cost.

This example illustrates how an analyst could use choice data to either reject LLR cost or to identify the information cost parameters $\left(\beta_{12}, \beta_{21}\right)$. In Proposition 11 in the online Appendix we formally show that choice probabilities are consistent with LLR cost if and only if a solution of the form (14) exists.

In general, when there are more than two states and actions the analyst might need data from multiple decision problems to point identify β. For a general decision problem the model admits $|\Theta|(|\Theta| - 1)$ degrees of freedom and (10) imposes $|\Theta| \times (1/2)|A| \times (|A| - 1)$ linear equations on β which suggests that to identify the analyst needs to observe behavior in

$$\frac{|\Theta| - 1}{\frac{1}{2}\,|A| \times \left(|A| - 1\right)}$$

decision problems. Given the data, solving numerically from the coefficients β is easy as the corresponding system of equations is linear.[23]

E. *Perception Tasks*

In this section we study the implications of the LLR cost function for perception tasks, a well known and long studied family of decision problems. In a perception

[23] Due to the linear structure of the implied restrictions, one could also construct finite sample tests for the LLR model using standard econometric methods, but this is beyond the scope of this paper.

task an agent is shown an even number of dots, with each dot either red or blue. The agent guesses whether there are more blue or red dots, and get rewarded if they guess correctly.

First, the total number n of dots is fixed. Then, subjects are told the value of n, and that the number i of red dots will be drawn uniformly from the set $\Theta = \{0, \ldots, n/2 - 1, n/2 + 1, \ldots, n\}$. The state where the number of blue and red dots is exactly equal to $n/2$ is ruled out to simplify the exposition. The set of actions is $A = \{R, B\}$ and the utility function is

$$u(a,i) = \begin{cases} 1 & \text{if} \quad a = B \quad \text{and} \quad i > n/2 \\ 1 & \text{if} \quad a = R \quad \text{and} \quad i < n/2 \\ 0 & \text{otherwise.} \end{cases}$$

Such perception tasks can be used to model many applied learning problems. For example, each dot could correspond to a voter whose color indicates whether they vote for the red or blue party and the agent is an analyst trying to predict which party will obtain the majority of votes in the election.[24] In a typical experiment subjects observe 100 dots each of which is either red or blue on a screen (see, e.g., Caplin and Dean 2013; Dean and Neligh 2017) and are asked whether there are more red or blue dots.

As in the case of binary decision problems, it is without loss of generality to assume that $\mu_i(B)$ is strictly between zero and one in every state. For a vector of distributions over actions (μ_i), the decision-maker guesses correctly in state i with probability

$$m_i = \begin{cases} \mu_i(B) & \text{if} \quad i > n/2 \\ \mu_i(R) & \text{if} \quad i < n/2. \end{cases}$$

Intuitively, it should be harder to guess correctly when the difference in the number of dots of different colors is small, i.e., when i is close to $n/2$. For example, it should be harder to predict the winner in a close election than in an election where one of the candidates has a large lead. Also, it is a well-established fact in the psychology,[25] neuroscience,[26] and economics[27] literatures that so called *psychometric functions*—the relation between the strength of a stimulus offered to a subject and the probability that the subject identifies this stimulus—are sigmoidal (i.e., S-shaped), so that the probability that a subject chooses B transitions smoothly from values close to zero to values close to one when the number of blue dots increases.

As Dean and Neligh (2017) note, under mutual information cost (and a uniform prior, as in the experimental setup described above), the optimal experiment $\mu^\star$

[24] Polling provides an interesting example of flexible information acquisition. Even if the only basic experiment available to a pollster is to call a voter and ask for her opinion, practically any experiment can be constructed as a compound experiment by deciding when to stop polling; i.e., the pollster with prior p can choose (perhaps at random) thresholds $p_1 < p < p_2$ and keep polling until her posterior reaches either p_1 or p_2. See Morris and Strack (2018) for a formalization of this idea.

[25] See, e.g., chapter 7 in Green and Swets (1966) or chapter 4 in Gescheider (1997).

[26] E.g., Krajbich, Armel, and Rangel (2010); Tavares, Perona, and Rangel (2017).

[27] See, e.g., Mosteller and Nogee (1951).

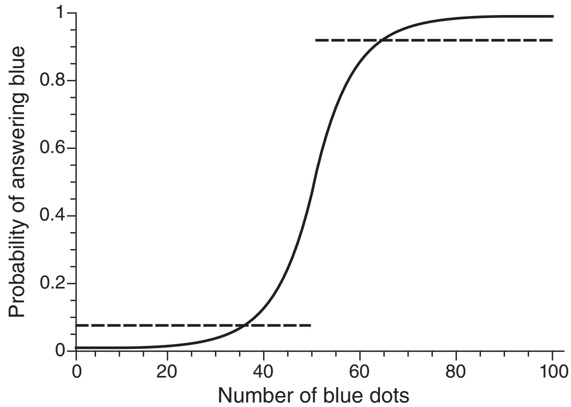

FIGURE 3

Note: Predicted probability of guessing that there are more blue dots as a function of the state, for the LLR cost with $\beta_{ij} = 1/(i-j)^2$ (solid line) and for mutual information cost (dashed line).

must induce a probability of guessing correctly that is state independent.[28] As shown by Matějka and McKay (2015), conditional on a state i, the log-likelihood ratio $\log(\mu_i(B)/\mu_i(R))$ between the two actions must equal the difference in payoffs $u(B,i) - u(R,i)$, up to a constant. Hence, the probability of a correct choice must be the same for any two states that lead to the same utility function over actions, such as the state in which there are 51 blue dots out of 100 and the state in which there are 99 blue dots.

As this is a one-dimensional information acquisition problem, we can apply the specification $\beta_{ij} = \kappa/(i-j)^2$ of the LLR cost. As can be seen in Figure 3, this LLR cost predicts a sigmoidal relation between the state and the choice probability. Thus, the model matches the qualitative features of choice probabilities commonly observed in practice. Of course, this could be similarly achieved using other cost functions that take into account the difficulty of distinguishing between similar states, such as the neighborhood cost function introduced by Hébert and Woodford (2020).

To gain additional insight, we now consider a more basic assumption on the cost function. Rather than assuming a particular specification, we assume that the coefficients (β_{ij}) are *strictly decreasing in the distance between states*: There exists a positive and strictly decreasing function f such that $\beta_{ij} = f(|i-j|)$ for all pairs of states. The condition captures the idea that states that are closer to each other are harder to distinguish. Even under this general nonparametric assumption, the LLR cost function leads to the intuitive prediction that the decision-maker will guess correctly with strictly higher probability when the difference in the number of dots of different colors is smaller:

[28] It is well known that under mutual information costs the physical features of the states (such as distance or similarity) do not affect the cost of information acquisition (see, e.g., Mackowiak, Matějka, and Wiederholt 2018).

PROPOSITION 6: *Consider the perception task above. Let C be an LLR cost function where the parameters* (β_{ij}) *are strictly increasing in the distance between states. Then, the resulting optimal probabilities* (m_i) *of guessing correctly satisfy* $m_i > m_j$ *whenever* $|i - (n/2)| > |j - (n/2)|$.

F. *The Effect of Greater Incentives*

We now apply the characterization of Proposition 3 to study in more detail the classic problem of predicting the probability of choosing between two options as a function of their relative values. In its simplest implementation, it consists of a task where there are two equally likely states, a subject must choose between two actions a_1 and a_2, and each action yields a reward with payoff $v \in \mathbb{R}$ when chosen in the corresponding state, and zero otherwise. Compared to Section VD, we focus here on the question of how the decision-maker's behavior varies as a function of v.

In order to interpret changes in the parameter v it is necessary to fix a cardinal representation of payoffs and to define an interval of possible values for v. If the decision-maker is risk neutral and rewards are monetary, then v can represent the amount paid to the subject. If the decision-maker is risk averse or her risk attitudes are unknown, then subjects can be paid using probabilistic prizes.[29]

The next result derives the optimal choice probabilities in a binary choice problem under a symmetric LLR cost function. Without loss of generality we restrict our attention to choice probabilities where both actions are chosen with strictly positive probability in every state. The result follows by rearranging the optimality conditions of Proposition 3.

PROPOSITION 7: *In a binary choice problem, let* $\mu_i[v]$ *denote the optimal choice probability of choosing action* a_i, *in state* i, *as a function of the reward* v, *under an LLR cost function. Assume the cost function satisfies* $\beta_{12} = \beta_{21} = \beta$. *Then* $\mu_1[v] = \mu_2[v] = m[v]$, *where*

$$m[v] = \frac{e^{\eta\left(\frac{v}{2\beta}\right)}}{1 + e^{\eta\left(\frac{v}{2\beta}\right)}},$$

and $\eta: \mathbb{R} \to \mathbb{R}$ *is the inverse of the function* $x \mapsto 2x + e^x - e^{-x}$.

As shown in Figure 4, and as can be easily proved analytically, the optimal choice probabilities $\mu[v]$ are a sigmoidal function of the payoff v. The prediction is in line with other standard models that involve noise or unobserved heterogeneity, including mutual information cost. Indeed, as shown by Matějka and McKay

[29] To illustrate, let x and y be two monetary prizes, with $x > y$. We continue to assume that the decision-maker's preferences are consistent with expected utility, and normalize, without loss of generality, their utility function to assign utility 1 to x and utility -1 to y. A lottery that delivers x with probability p and y with probability $1 - p$ has expected utility $2p - 1$. We define each payoff v in the interval $[-1, 1]$ as the expected utility of such a lottery. This approach is well known in the implementation of scoring rules, where it allows to reward a decision-maker using a linear payoff, and circumvents the need of eliciting the decision-maker's degree of risk aversion (see, for example, Lambert 2018; Sandroni and Shmaya 2013, and the references therein). The same approach has been more recently applied in rational inattention by Caplin et al. (2020).

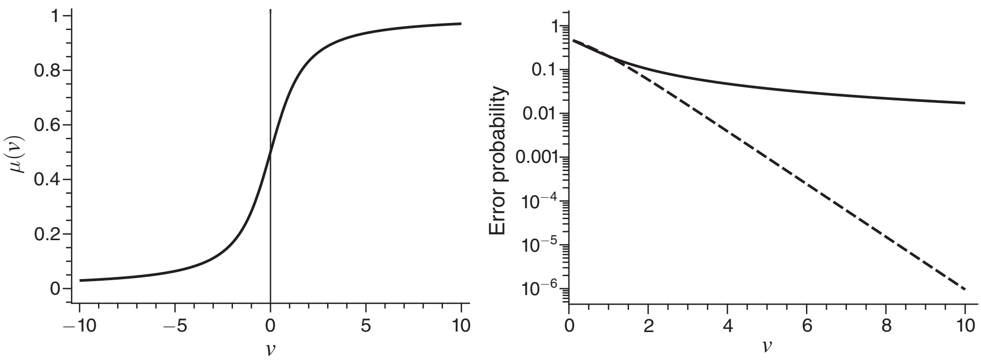

FIGURE 4

Notes: On the left: the optimal choice probabilities in a binary decision problem for an LLR cost function. On the right: the implied probabilities of choosing incorrectly at different levels of incentives v if the agent chooses correctly with 80 percent probability for $v = 1$ for the LLR cost (solid line) and mutual information cost (dashed line) on a log scale.

(2015), under mutual information the optimal choice probabilities follow a logistic relation, where the probability of matching the state, as a function of v, is given by

$$\frac{e^{\frac{v}{2\lambda}}}{1 + e^{\frac{v}{2\lambda}}},$$

and $\lambda > 0$ is the parameter controlling the cost of information acquisition. The two functional forms are similar, with the only difference being the transformation η. The function is strictly increasing and S-shaped, onto, and satisfies $\eta(x) = \eta(-x)$ (and hence $\eta(0) = 0$).[30]

While both the LLR and the mutual information models lead to choice probabilities that are sigmoidal, the two theories lead to different predictions on how the probabilities of errors scale with the payoff v. Figure 4 displays the implied probabilities with which a decision-maker takes a correct choice as a function of v, under the two theories. To make the comparison meaningful, the parameters β and λ are chosen so that in both models the agent chooses incorrectly with probability 20 percent when the payoff is $v = 1$.

As one can see in the figure, the probability of choosing correctly reacts more strongly to incentives under mutual information cost. For example, suppose that the payoff v is measured in dollars. A simple calculation shows that under mutual information cost, if the decision-maker chooses incorrectly with probability 20 percent

[30] For the two models to be distinguished empirically, it is necessary to isolate the nonlinearity described by η from other confounding effects. This can be difficult when v represents dollar amounts, as the same choice probabilities obtained under log-likelihood ratio and risk neutrality would obtain under mutual information and a utility function η over money. This is however not an issue if payoffs are defined using probabilistic prizes and preferences are consistent with expected utility. Allowing for more general preferences can lead to new difficulties. For example, the same choice probabilities we obtain with LLR cost can be obtained under mutual information when the decision-maker has nontrivial attitudes towards how lotteries resolve over time, captured by the curvature of η. For preferences beyond expected utility, Caradonna (2021) provides a methodology for obtaining quasi-linear representations which could be used to extend our approach.

when $v = \$1$, then she must choose incorrectly with probability less than one in a million if $v = \$10$. LLR costs imply that this probability is about $1/60$.[31] These are starkly different predictions about behavior which can be tested experimentally.

The finding is not special to this example. Under logistic choice (e.g., as in Matějka and McKay 2015), the probability of making a mistake decays quickly, as v grows, at the exponential rate e^{-v}. Under the LLR cost function the same probability decreases at the much slower rate $1/v$. This follows from Proposition 7, together with the fact that as v increases, the transformation η approximates the logarithm.

VI. Bayesian LLR Cost

Given a prior q and an LLR cost function C, one can express the cost of an experiment μ in terms of the distribution π_μ of the posterior belief $p \in \Delta(\Theta)$ that it induces, via

$$C(\mu) = \int F(p) - F(q) \mathrm{d}\pi_\mu(p), \tag{15}$$

where

$$F(p) = \sum_{i,j} \beta_{ij} \frac{p_i}{q_i} \log\left(\frac{p_i}{p_j}\right).$$

This follows from the definition of the LLR cost, together with Bayes' law, which states that given a prior q and a signal s, the posterior p is given by $\log(p_i/p_j) = \log(q_i/q_j) + \log(d\mu_i/d\mu_j)(s)$. This reformulation shows that the LLR cost is posterior separable (Caplin and Dean 2013).

A stronger property studied in the literature is *uniform* posterior separability, where the function F is independent of the prior q. In addition to being standard, this assumption ensures, for instance, that in a dynamic environment an agent is indifferent between performing two experiments—with the choice of the second one perhaps depending on the outcome of the first—and carrying out the Blackwell equivalent one-shot experiment.

As we now show, this assumption can be accommodated in our framework by allowing the cost $C(\mu, q)$ of an experiment μ to be a function the prior, where for each prior the cost function $C(\cdot, q)$ belongs to the LLR family, and the resulting coefficients $(\beta_{ij}(q))$ depend on the prior. While any functional relation between the prior and the coefficients is consistent with LLR cost, there is a unique choice that makes the Bayesian LLR cost function uniform posterior separable, as the next proposition shows. An analogous result was derived independently by Bloedel and Zhong (2020).

[31] For an alternative interpretation, suppose the decision-maker is paid in chance rather than money, so that the payoff v denotes the probability of receiving a prize conditional on making a correct choice. Suppose that when v is 0.05 percent, the decision-maker makes a mistake with probability 20 percent. Then, if the probability v is increased to 0.5 percent, the prediction under mutual information is that the decision-maker must make a mistake with a probability that is less than one in a million. Under log-likelihood ratio the probability is about $1/60$.

PROPOSITION 8: *A Bayesian LLR cost function C given by*

$$C(\mu, q) = \sum_{i,j} \beta_{ij}(q) D_{\mathrm{KL}}(\mu_i \| \mu_j)$$

is uniform posterior separable if and only if there exist positive constants $(b_{ij})_{i,j \in \Theta, i \neq j}$ *such that for all priors* $q \in \mathcal{P}(\Theta)$ *with full support,* $\beta_{ij}(q) = b_{ij} q_i$.

Both prior independence and constant marginal costs are reasonable assumptions when modeling common actions of information acquisitions, such as performing a measurement or drawing samples. A first implication of Proposition 8 is that the two assumptions are incompatible with uniform posterior separability, a desirable property in a dynamic setting. This is discussed in depth by Bloedel and Zhong (2020).

Proposition 8 also shows that uniform posterior separability is possible if the coefficients β_{ij} are allowed to change with the prior. Letting

$$F(p) = \sum_{i,j} b_{ij} p_i \log\left(\frac{p_i}{p_j}\right),$$

and substituting this into (15), we see that Bayesian LLR cost of an experiment can be represented as the expected change of F from the prior q to the posterior p induced by the signal, for a fixed choice of (b_{ij}). That is, the cost of the experiment equals

$$C(\mu, q) = \int \left[F(p) - F(q)\right] \mathrm{d}\pi(p), \tag{16}$$

and in particular it is uniform posterior separable. For a given prior, this cost is the LLR cost with $\beta_{ij} = b_{ij} q_i$, so that, in terms of the distributions (μ_i), this cost is

$$C(\mu, q) = \sum_{i,j} b_{ij} q_i D_{\mathrm{KL}}(\mu_i \| \mu_j). \tag{17}$$

Proposition 8 implies that the only uniform posterior separable LLR cost potentially assigns different cost to the same experiment at different prior beliefs. Nevertheless, some experiments may be assigned a cost that does not depend on the prior.

VII. Verification and Falsification

All the specifications of the LLR cost we have discussed in the previous sections have the property that the coefficients are symmetric across states, so that $\beta_{ij} = \beta_{ji}$. In this section we explain why some information costs are best modeled by specifications that break this symmetry.

It is well understood that verification and falsification are fundamentally different forms of empirical research. This can be seen most clearly through Karl Popper's famous example of the statement "all swans are white." Regardless of how many white swans are observed, no amount of evidence can imply that the next one will be white. However, observing a single black swan is enough to prove the statement false. Popper's argument highlights a crucial asymmetry between verification and falsification. A given experiment, such as the observation of swans, can make it feasible to reject a hypothesis, yet have no power to prove that the same hypothesis is true.

Table 1—The Set of States is $\Theta = \{a, e\}$.

	Panel A. Experiment I			Panel B. Experiment II	
	s_1	s_2		s_1	s_2
a	$1-\varepsilon^2$	ε^2	a	$1-\varepsilon$	ε
e	$1-\varepsilon$	ε	e	$1-\varepsilon^2$	ε^2

Notes: In both experiments $S = \{s_1, s_2\}$. Under experiment I, observing the signal realization s_2 rejects the hypothesis that the state is a (up to a small probability of error ε^2). Under experiment II, observing s_2 verifies the same hypothesis.

This principle extends from science to everyday life. In a legal case, the type of evidence necessary to prove that a person is guilty can be quite different from the type of evidence necessary to demonstrate that a person is innocent. In a similar way, corroborating the claim "Ann has a sibling" might require empirical evidence (such as the outcome of a DNA test) that is distinct from the sort of evidence necessary to prove that she has no siblings.

In this section we show that the asymmetry between verification and falsification can be captured by the LLR cost. As an example, we consider a state space $\Theta = \{a, e\}$ that consists of two hypotheses. For simplicity, let $\{a\}$ correspond to the hypothesis "all swans are white" and $\{e\}$ the complementary hypothesis "there exists a nonwhite swan." Imagine a decision-maker who attaches equal probability to the each state, and consider the experiments described in Table 1:[32]

- In experiment I, regardless of the state, an uninformative signal realization s_1 occurs with probability greater than $1 - \varepsilon$, where ε is positive and small. If a nonwhite swan exists, then one is observed with probability ε. Formally, this corresponds to observing the signal realization s_2. If all swans are white, then signal s_1 is observed, up to a minuscule probability of error ε^2. Hence, conditional on observing s_2, the decision-maker's belief in state a approaches zero, while conditional on observing s_1 the decision-maker's belief remains close to the prior. So, the experiment can reject the hypothesis that the state is a, but cannot verify it. We set the probability of observing a nonwhite swan in state a equal to ε^2 rather than zero, to ensure that log-likelihood ratios are finite for each observation, and hence that the experiment has finite cost.

- In experiment II the roles of the two states are reversed: if all swans are white, then this fact is revealed to the decision-maker with probability ε. If there is a nonwhite swan, then the uninformative signal s_1 is observed (up to the small probability of error ε^2). Conditional on observing s_2, the decision-maker's belief in state a approaches one, and conditional on observing s_1 the decision-maker's belief is essentially unchanged. Thus, the experiment can verify the hypothesis that the state is a, but cannot reject it.

[32] Popper (1959) intended verification and falsifications as deterministic procedures, which exclude even small probabilities of error. In our informal discussion we do not distinguish between events that are deemed extremely unlikely (such as thinking of having observed a black swan in world where all swans are white) and events that have zero probability. In their work on falsifiability, Olszewski and Sandroni (2011) ascribe to Cournot (1843) the idea that unlikely events must be treated as impossible.

As shown by the example, permuting the state-dependent distributions of an experiment may affect its power to verify or falsify an hypothesis. However, permuting the role of the states may, in reality, correspond to a completely different type of empirical investigation. For instance, experiment I can be easily implemented in practice: as an extreme example, the decision-maker may look up in the sky. There is a small chance a nonwhite swan will be observed; if not, the decision-maker's belief will not change by much. It is not obvious exactly what tests or samples would be necessary to implement experiment II, which must be able to reveal that all swans are white, let alone to conclude that the two experiments should be equally costly.

We conclude that in order for a model of information acquisition to capture the difference between verification and falsification, the cost of an experiment should not necessarily be invariant with respect to a permutation of the states. In our model, this can be captured by assuming that the coefficients (β_{ij}) are nonsymmetric, i.e., that β_{ij} and β_{ji} are not necessarily equal. For instance, the cost of experiments I and II in Table 1 will differ whenever the coefficients of the LLR cost satisfy $\beta_{ae} \neq \beta_{ea}$. For example, set $\beta_{ae} = \kappa$ and $\beta_{ea} = 0$, and consider small ε. Then, to first order in ε, the cost of experiment I is $\kappa\varepsilon$, while the cost of experiment II is a factor of $\log(1/\varepsilon)$ higher. Hence the ratio between the costs of these experiments is arbitrarily high for small ε.

VIII. Related Literature

The question of how to quantify the amount of information provided by an experiment is the subject of a long-standing and interdisciplinary literature. Kullback and Leibler (1951) introduced the notion of Kullback-Leibler divergence as a measure of distance between statistical populations. Kelly (1956); Lindley (1956); Marschak (1959); and Arrow (1972) apply mutual information to the problem of ordering information structures.

More recently, Hansen and Sargent (2001) and Strzalecki (2011) adopted KL-divergence as a tool to model robust decision criteria under uncertainty. Cabrales, Gossner, and Serrano (2013) derive Shannon entropy as an index of informativeness for experiments in the context of portfolio choice problems (see also Cabrales, Gossner, and Serrano 2017). Frankel and Kamenica (2018) put forward an axiomatic framework for quantifying the value and the amount of information in an experiment.

Rational Inattention.—As discussed in the introduction, our work is also motivated by the recent literature on rational inattention. A complete survey of this area is beyond the scope of this paper; we instead refer the interested reader to Caplin (2016) and Mackowiak, Matějka, and Wiederholt (2018) for perspectives on this growing literature.

Decision Theory.—Our axiomatic approach differs both in terms of motivation and techniques from other results in the literature. Caplin and Dean (2015) study the revealed preference implications of rational inattention models, taking as a primitive state-dependent random choice data. Within the same framework, Caplin, Dean, and Leahy (2018) characterize mutual information cost, Chambers, Liu, and Rehbeck

(2017) study nonseparable models of costly information acquisition, and Denti (2022) provides a revealed preference of posterior separability. Decision theoretic foundations for models of information acquisition have been studied by de Oliveira (2014); De Oliveira et al. (2017); and Ellis (2018). Mensch (2018) provides an axiomatic characterization of posterior separable cost functions.

The Wald Model of Sequential Sampling.—The notion of constant marginal costs over independent experiments goes back to Wald's (1945) classic sequential sampling model; our axioms extend some of Wald's ideas to a model of flexible information acquisition. In its most general form, Wald's model considers a decision-maker who acquires information by collecting multiple independent copies of a fixed experiment, and incurs a cost equal to the number of repetitions. In this model, every stopping strategy corresponds to an experiment, and so every such model defines a cost over some family of experiments. It is easy to see that such a cost satisfies our axioms.

Morris and Strack (2018) consider a continuous-time version where the decision-maker observes a one-dimensional diffusion process whose drift depends on the state, and incurs a cost proportional to the expected time spent observing. This cost is again easily seen to satisfy our axioms, and indeed, for the experiments that can be generated using this sampling process, they show that the expected cost of a given distribution over posteriors is of the form obtained in Proposition 2. One may view the result in Morris and Strack as complementary evidence that the cost function obtained in Proposition 2 is a natural choice for one-dimensional information acquisition problems.

Dynamic Information Acquisition Models.—Hébert and Woodford (2019, 2020); Zhong (2017, 2019); Morris and Strack (2018); and Bloedel and Zhong (2020) relate cost functions over experiments and sequential models of costly information acquisition. In these papers, the cost $C(\mu)$ is the minimum expected cost of generating the experiment μ by means of a dynamic sequential sampling strategy.

Hébert and Woodford (2020) propose and characterize a family of "neighborhood-based" cost functions that generalize mutual information, and allow for the cost of learning about states to be affected by their distance. In a perception task, these costs are flexible enough to accommodate optimal response probabilities that are S-shaped, similarly to our analysis in Section V. The LLR cost does not generalize mutual information, but has a structure similar to a neighborhood-based cost where the neighboring structure consists of all pairs of states.

Zhong (2017) and Bloedel and Zhong (2020) provide general conditions for a cost function over experiments to be induced by some dynamic model of information acquisition. Zhong (2019) studies a dynamic model of nonparametric information acquisition, where a decision-maker can choose any dynamic signal process as an information source, and pays a flow cost that is a function of the informativeness of the process. A key assumption is discounting of delayed payoffs. The paper shows that the optimal strategy corresponds to a Poisson experiment.

Information Theory.—This paper is also related to the axiomatic literature in information theory characterizing different notions of entropy and information measures. Ebanks, Sahoo, and Sander (1998) and Csiszár (2008) survey and summarize the literature in the field. In the special case where $|\Theta| = 2$ and the coefficients (β_{ij}) are set to 1, the function (1) is also known as *J-divergence*. Kannappan and Rathie (1988) provide an axiomatization of J-divergence, under axioms very different from the ones in this paper. A more general representation appears in Zanardo (2017).

Ebanks, Sahoo, and Sander (1998) characterize functions over tuples of measures with finite support. They show that a condition equivalent to our additivity axiom leads to a functional form similar to (1). Their analysis is however quite different from ours: their starting point is an assumption which, in the notation of this paper, states the existence of a map $F: \mathbb{R}^{\Theta} \to \mathbb{R}$ such that the cost of an experiment $(S, (\mu_i))$ with finite support takes the form $C(\mu) = \sum_{s \in S} F((\mu_i(s))_{i \in \Theta})$. This assumption of additive separability does not seem to have an obvious economic interpretation, nor to be related to our motivation of capturing constant marginal costs in information production.

Probability Theory.—The results in Mattner (1999, 2004) have, perhaps, the closest connection with this paper. Mattner studies functionals over the space probability measures over $\mathbb{R}$ that are additive with respect to convolution. As we explain in the next section, additivity with respect to convolution is a property that is closely related to Axiom 2. We draw inspiration from Mattner (1999) in applying the study of cumulants to the proof of Theorem 1. However, the difference in domain makes the techniques in Mattner (1999, 2004) not applicable to this paper.

IX. Proof Sketch

In this section we informally describe some of the ideas involved in the proof of Theorem 1. We consider the binary case where $\Theta = \{0, 1\}$ and so there is only one relevant log-likelihood ratio $\ell = \ell_{10}$. The proof of the general case is more involved, but conceptually similar.

Step 1: Let C satisfy Axioms 1–4. Conditional on each state i, an experiment μ induces a distribution σ_i for ℓ. Two experiments that induce the same pair of distributions (σ_0, σ_1) are equivalent in the Blackwell order. Thus, by Axiom 1, C can be identified with a map $c(\sigma_0, \sigma_1)$ defined over all pairs of distributions induced by some experiment μ.

Step 2: Axioms 2 and 3 translate into the following properties of c. The product $\mu \otimes \nu$ of two experiments induces, conditional on i, a distribution for ℓ that is the convolution of the distributions induced by the two experiments.[33] Axiom 2 is equivalent to c being additive with respect to convolution, i.e.,

$$c(\sigma_0 * \tau_0, \sigma_1 * \tau_1) = c(\sigma_0, \sigma_1) + c(\tau_0, \tau_1).$$

[33] Recall that given two distributions σ and ν over $\mathbb{R}$, their convolution is the distribution of the random variable $X + Y$, where X is a random variable distributed according to σ, Y according to ν, and the two random variables are independent. When two experiments are independent (in the sense described in Section I), their log-likelihood ratios are independent random variables conditional on the state. The crucial observation is that the log-likelihood ratio of the product experiment is the sum of the individual log-likelihood ratios, and thus its distribution conditional on the state is the convolution of theirs.

Axiom 3 is equivalent to c satisfying for all $\alpha \in [0,1]$,

$$c\big(\alpha\sigma_0 + (1-\alpha)\delta_0, \alpha\sigma_1 + (1-\alpha)\delta_0\big) = \alpha c(\sigma_0,\sigma_1),$$

where δ_0 is the degenerate measure at 0. Axiom 4 translates into continuity of c with respect to total variation and the first N moments of σ_0 and σ_1.

Step 3: As is well known, many properties of a probability distribution can be analyzed by studying its moments. We apply this idea to the study of experiments, and show that under our axioms the cost $c(\sigma_0,\sigma_1)$ is a function of the first N moments of the two measures, for some (arbitrarily large) N. Given an experiment μ, we consider the experiment

$$\frac{1}{n}\cdot(\mu \otimes \cdots \otimes \mu),$$

in which with probability $(n-1)/n$ no information is produced, and with the remaining probability the experiment μ is carried out n times. By Axioms 2 and 3, the cost of this experiment is equal to the cost of μ.[34] We show that these properties, together with the continuity axiom, imply that the cost of an experiment is a function G of the moments of (σ_0,σ_1):

$$c(\sigma_0,\sigma_1) = G\big[m_{\sigma_0}(1),\ldots,m_{\sigma_0}(N), m_{\sigma_1}(1),\ldots,m_{\sigma_1}(N)\big], \tag{18}$$

where $m_{\sigma_i}(n)$ is the nth moment of σ_i. Each $m_{\sigma_i}(n)$ is affine in σ_i, hence Step 2 implies that G is affine with respect to mixtures with the zero vector.

Step 4: It will be useful to analyze a distribution not only through its moments but also through its cumulants. The nth cumulant $\kappa_\sigma(n)$ of a probability measure σ is the nth derivative at zero of the logarithm of its characteristic function. By a combinatorial characterization due to Leonov and Shiryaev (1959), $\kappa_\sigma(n)$ is a polynomial function of the first n moments $m_\sigma(1),\ldots,m_\sigma(n)$. For example, the first cumulant is the expectation $\kappa_\sigma(1) = m_\sigma(1)$, the second is the variance, and the third is $\kappa_\sigma(3) = m_\sigma(3) - 2m_\sigma(2)m_\sigma(1) + 2m_\sigma(1)^3$. Step 3 and the result by Leonov and Shiryaev (1959) imply that the cost of an experiment is a function H of the cumulants of (σ_0,σ_1):

$$c(\sigma_0,\sigma_1) = H\big[\kappa_{\sigma_0}(1),\ldots,\kappa_{\sigma_0}(N),\kappa_{\sigma_1}(1),\ldots,\kappa_{\sigma_1}(N)\big], \tag{19}$$

where $\kappa_{\sigma_i}(n)$ is the nth cumulant of σ_i.

Step 5: Cumulants satisfy a crucial property: the cumulant of a sum of two independent random variables is the sum of their cumulants. So, they are additive with respect to convolution. By Step 2, this implies that H is additive. We show that H is in fact a linear function. This step is reminiscent of the classic Cauchy equation

[34] For n large, this experiment has a very simple structure: With high probability it is uninformative, and with probability $1/n$ is highly revealing about the states.

problem. That is, understanding under what conditions a function $\phi\colon \mathbb{R} \to \mathbb{R}$ that satisfies $\phi(x+y) = \phi(x) + \phi(y)$ must be linear. In Theorem 4 we show, very generally, that any additive function from a subset $\mathcal{K} \subset \mathbb{R}^d$ to $\mathbb{R}_+$ is linear, provided $\mathcal{K}$ is closed under addition and has a nonempty interior. We then proceed to show that both of these conditions are satisfied if $\mathcal{K}$ is taken to be the domain of H, and thus deduce that H is linear.

Step 6: In the last step we study the implications of (18) and (19). We apply the characterization by Leonov and Shiryaev (1959) and show that the affinity with respect to the origin of the map G, and the linearity of H, imply that H must be a function solely of the first cumulants $\kappa_{\sigma_0}(1)$ and $\kappa_{\sigma_1}(1)$. That is, C must be a weighted sum of the expectations of the log-likelihood ratio ℓ conditional on each state.

X. Conclusions

We put forward an axiomatic approach to modeling the cost of information acquisition, characterizing a family of cost functions that capture a notion of constant marginal costs in the production of information. We propose a number of possible avenues for future research, all of which would require the solution of some nontrivial technical challenges: The first is an extension of our framework beyond the setting of a finite set of states to a continuum of states. This is natural in the context of one-dimensional problems. Second, one could consider multidimensional problems in which Θ is a finite subset of $\mathbb{R}^d$, and study a generalization of the one-dimensional functional form we obtain in Section III. Third, there are a number of settings which have been modeled using mutual information cost, where it may be of interest to understand the sensitivity of the conclusions to this assumption (see, e.g., Van Nieuwerburgh and Veldkamp 2010). Finally, a possible definition for convex cost functions over experiments is given by the supremum over a family of LLR costs. It may be interesting to understand if such costs are characterized by simple axioms.

REFERENCES

Arrow, Kenneth J. 1972. "The Value of and Demand for Information." In *Decision and Organization: A Volume in Honor of Jacob Marschak*, edited by C. B. McGuire and Roy Radner, 131–39. Amsterdam: Elsevier.

Arrow, Kenneth J. 1985. "Informational Structure of the Firm." *American Economic Review* 75 (2): 303–07.

Arrow, Kenneth J, David Blackwell, and Meyer A. Girshick. 1949. "Bayes and Minimax Solutions of Sequential Decision Problems." *Econometrica* 17 (3–4): 213–44.

Austin, Tim D. 2006. "Entropy and Sinai Theorem." Unpublished.

Blackwell, David. 1951. "Comparison of Experiments." In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, edited by Jerzy Neyman, 93–102. Berkeley, CA: University of California Press.

Bloedel, Alexander W., and Weijie Zhong. 2020. "The Cost of Optimally-Acquired Information." Unpublished.

Bohnenblust, Henri Frédéric, Lloyd S. Shapley, and Seymour Sherman. 1949. *Reconnaissance in Game Theory*. Santa Monica, CA: Rand Corporation.

Borwein, Jonathan M., and Jon D. Vanderwerff. 2010. *Convex Functions: Constructions, Characterizations and Counterexamples*, Vol. 109. Cambridge, UK: Cambridge University Press.

Brouwer, Luitzen. 1911. "Beweis der invarianz des n-dimensionalen gebiets." *Mathematische Annalen* 71 (3): 305–13.
Cabrales, Antonio, Olivier Gossner, and Roberto Serrano. 2013. "Entropy and the Value of Information for Investors." *American Economic Review* 103 (1): 360–77.
Cabrales, Antonio, Olivier Gossner, and Roberto Serrano. 2017. "A Normalized Value for Information Purchases." *Journal of Economic Theory* 170: 266–88.
Caplin, Andrew. 2016. "Measuring and Modeling Attention." *Annual Review of Economics* 8: 379–403.
Caplin, Andrew, and Mark Dean. 2013. "Behavioral Implications of Rational Inattention with Shannon Entropy." NBER Working Paper 19318.
Caplin, Andrew, and Mark Dean. 2015. "Revealed Preference, Rational Inattention, and Costly Information Acquisition." *American Economic Review* 105 (7): 2183–2203.
Caplin, Andrew, Dániel Csaba, John Leahy, and Oded Nov. 2020. "Rational Inattention, Competitive Supply, and Psychometrics." *Quarterly Journal of Economics* 135 (3): 1681–1724.
Caplin, Andrew, Mark Dean, and John Leahy. 2016. "Rational Inattention, Optimal Consideration Sets and Stochastic Choice." Unpublished.
Caplin, Andrew, Mark Dean, and John Leahy. 2018. "Rational Inattentive Behavior: Characterizing and Generalizing Shannon Entropy." NBER Working Paper 23652.
Caradonna, Peter P. 2021. "Preference Regression." Unpublished.
Chambers, Christopher P., Ce Liu, and John Rehbeck. 2017. "Nonseparable Costly Information Acquisition and Revealed Preference." Unpublished.
Chan, Jimmy, Alessandro Lizzeri, Wing Suen, and Leeat Yariv. 2017. "Deliberating Collective Decisions." *Review of Economic Studies* 85 (2): 929–63.
Christie, Agatha. 1934. *Murder on the Orient Express.* London: Collins Crime Club.
Cournot, Antoine Augustin. 1843. *Exposition de la théorie des chances et des probabilités.* Paris, France: Hachette Livre-BNF.
Cover, Thomas M., and Joy A. Thomas. 2012. *Elements of Information Theory.* Hoboken, NJ: John Wiley & Sons.
Csiszár, Imre. 2008. "Axiomatic Characterizations of Information Measures." *Entropy* 10 (3): 261–73.
Dean, Mark, and Nathaniel Neligh. 2017. "Experimental Tests of Rational Inattention." Unpublished.
Denti, Tommaso. 2022. "Posterior Separable Cost of Information." *American Economic Review* 112 (10): 3215–59.
Denti, Tommaso, Massimo Marinacci, and Luigi Montrucchio. 2020. "A Note on Rational Inattention and Rate Distortion Theory." *Decisions in Economics and Finance* 43 (1): 75–89.
de Oliveira, Henrique. 2014. "Axiomatic Foundations for Entropic Costs of Attention." Unpublished.
de Oliveira, Henrique, Tommaso Denti, Maximilian Mihm, and Kemal Ozbek. 2017. "Rationally Inattentive Preferences and Hidden Information Costs." *Theoretical Economics* 12 (2): 621–54.
Dupuis, Paul, and Richard S. Ellis. 2011. *A Weak Convergence Approach to the Theory of Large Deviations*, Vol. 902. Hoboken, NJ: John Wiley & Sons.
Dvoretzky, Aryeh, Jack Kiefer, and Jacob Wolfowitz. 1953. "Sequential Decision Problems for Processes with Continuous Time Parameter. Testing Hypotheses." *Annals of Mathematical Statistics* 24 (2): 254–64.
Ebanks, Bruce, Prasanna Sahoo, and Wolfgang Sander. 1998. *Characterizations of Information Measures.* Singapore: World Scientific.
Ellis, Andrew. 2018. "Foundations for Optimal Inattention." *Journal of Economic Theory* 173: 56–94.
Frankel, Alexander, and Emir Kamenica. 2018. "Quantifying Information and Uncertainty." Unpublished.
Gescheider, George A. 1997. *Psychophysics: The Fundamentals.* 3rd ed. Oxfordshire, England: Psychology Press.
Green, David Marvin, and John Arthur Swets. 1966. *Signal Detection Theory and Psychophysics.* New York, NY: Wiley.
Hansen, LarsPeter, and Thomas J. Sargent. 2001. "Robust Control and Model Uncertainty." *American Economic Review* 91 (2): 60–66.
Hébert, Benjamin M., and Michael Woodford. 2019. "Rational Inattention When Decisions Take Time." NBER Working Paper 26415.
Hébert, Benjamin M., and Michael Woodford. 2020. "Neighborhood-Based Information Costs." NBER Working Paper 26743.
Jech, Thomas. 2013. *Set Theory.* Berlin, Germany: Springer Science & Business Media.
Kannappan, Pl., and P. N. Rathie. 1988. "An Axiomatic Characterization of J-divergence." In *Transactions of the Tenth Prague Conference on Information Theory, Statistical Decision Functions, Random Processes*, 29–36. New York: Springer.

Kelly, J. L., Jr. 1956. "A New Interpretation of Information Rate." *Bell System Technical Journal* 35 (4): 917–26.

Krajbich, Ian, Carrie Armel, and Antonio Rangel. 2010. "Visual Fixations and the Computation and Comparison of Value in Simple Choice." *Nature Neuroscience* 13 (10): 1292–98.

Kullback, Solomon, and Richard A. Leibler. 1951. "On Information and Sufficiency." *Annals of Mathematical Statistics* 22 (1): 79–86.

Lambert, Nicholas. 2018. "Probability Elicitation for Agents with Arbitrary Risk Preferences." Unpublished.

Le Cam, Lucien. 1996. "Comparison of Experiments: A Short Review." *Lecture Notes-Monograph Series* 30: 127–38.

Leonov, V. P., and Albert Nikolaevich Shiryaev. 1959. "On a Method of Calculation of Semi-invariants." *Theory of Probability and Its Applications* 4 (3): 319–29.

Liese, Friedrich, and Igor Vajda. 1987. *Convex Statistical Distances*, Vol. 95. Leipzig, Germany: Teubner.

Lindley, Dennis V. 1956. "On a Measure of the Information Provided by an Experiment." *Annals of Mathematical Statistics* 27 (4): 986–1005.

Mackowiak, Bartosz, Filip Matějka, and Mirko Wiederholt. 2018. "Rational Inattention: A Disciplined Behavioral Model." Unpublished.

Marschak, Jacob. 1959. "Remarks on the Economics of Information." Cowles Foundation Discussion Paper 293.

Matějka, Filip, and Alisdair McKay. 2015. "Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model." *American Economic Review* 105 (1): 272–98.

Mattner, Lutz. 1999. "What Are Cumulants?" *Documenta Mathematica* 4: 601–22.

Mattner, Lutz. 2004. "Cumulants Are Universal Homomorphisms into Hausdorff groups." *Probability Theory and Related Fields* 130 (2): 151–66.

Mensch, Jeffrey. 2018. "Cardinal Representations of Information." Unpublished.

Morris, Stephen, and Ming Yang. 2016. "Coordination and Continuous Choice." Unpublished.

Morris, Stephen, and Philipp Strack. 2018. "The Wald Problem and the Relation of Sequential Sampling and Static Information Costs." Unpublished.

Mosteller, Frederick, and Philip Nogee. 1951. "An Experimental Measurement of Utility." *Journal of Political Economy* 59 (5): 371–404.

Mu, Xiaosheng, Luciano Pomatto, Philipp Strack, and Omer Tamuz. 2021. "From Blackwell Dominance in Large Samples to Rényi Divergences and Back Again." *Econometrica* 89 (1): 475–506.

Olszewski, Wojciech, and Alvaro Sandroni. 2011. "Falsifiability." *American Economic Review* 101 (2): 788–818.

Popper, Karl. 1959. *The Logic of Scientific Discovery.* London: Routledge.

Sandroni, Alvaro, and Eran Shmaya. 2013. "Eliciting Beliefs by Paying in Chance." *Economic Theory Bulletin* 1 (1): 33–37.

Shiryaev, A. N. 1996. *Probability.* New York: Springer.

Sims, Christopher A. 2003. "Implications of Rational Inattention." *Journal of Monetary Economics* 50 (3): 665–90.

Sims, Christopher. 2010. "Rational Inattention and Monetary Economics." In *Handbook of Monetary Economics*, Vol. 3, edited by Benjamin M. Friedman and Michael Woodford, 155–181. Amsterdam: Elsevier.

Strzalecki, Tomasz. 2011. "Axiomatic Foundations of Multiplier Preferences." *Econometrica* 79 (1): 47–73.

Tao, Terence. 2011. "Brouwer's Fixed Point and Invariance of Domain Theorems, and Hilbert's Fifth Problem." Unpublished.

Tavares, Gabriela, Pietro Perona, and Antonio Rangel. 2017. "The Attentional Drift Diffusion Model of Simple Perceptual Decision-Making." *Frontiers in Neuroscience* 11: 468.

Van Nieuwerburgh, Stijn, and Laura Veldkamp. 2010. "Information Acquisition and Under-Diversification." *Review of Economic Studies* 77 (2): 779–805.

Wald, Abraham. 1945. "Sequential Tests of Statistical Hypotheses." *Annals of Mathematical Statistics* 16 (2): 117–86.

Wilson, Robert. 1975. "Informational Economies of Scale." *Bell Journal of Economics* 6 (1): 184–95.

Zanardo, Enrico. 2017. "How to Measure Disagreement." Unpublished.

Zhong, Weijie. 2017. "Indirect Information Measure and Dynamic Learning." Unpublished.

Zhong, Weijie. 2019. "Optimal Dynamic Information Acquisition." Unpublished.