NBER WORKING PAPER SERIES

THE ECONOMICS OF INFORMATION IN A WORLD OF DISINFORMATION: A SURVEY PART 2: DIRECT COMMUNICATION

Joseph E. Stiglitz
Andrew Kosenko

Working Paper 32050
http://www.nber.org/papers/w32050

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2024

We are grateful to Daphne Raban and Julia Włodarczuk, for their work on improving the paper, to Andrea Gurwitt, for editing, to Filippo Pavesi and Massimo Scotti, and to the many collaborators that worked with us over the years. Joseph Stiglitz gratefully acknowledges financial support from the Institute for New Economic Thinking. A slightly different version of this paper has been prepared for the Elgar Companion on Information Economics. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

The Economics of Information in a World of Disinformation: A Survey Part 2: Direct Communication
Joseph E. Stiglitz and Andrew Kosenko
NBER Working Paper No. 32050
January 2024
JEL No. D82,D86,D9

ABSTRACT

The paper surveys the recent work on economics of information with endogenous information structures where individuals can directly communicate information with each other. We consider the theoretical work on cheap talk, Bayesian persuasion, and information design, and review the implications of information control and information abundance for mis and disinformation. The relationship between information and market power is particularly important when social media can amplify and maintain harmful fictions that lead to polarization and undermine not only markets, but democratic discourse. We review both the "rational" decision-making paradigm, as well as departures from it, such as cases where decision makers can choose what to know, can allocate their attention in different ways or have behavioral biases that influence their information processing. We note some important connections to legal and media studies and highlight key messages in nontechnical language.

Joseph E. Stiglitz
5th Floor, Kravis Hall
Columbia University
665 West 130th St.
New York, NY 10027
and NBER
jes322@columbia.edu

Andrew Kosenko
Marist College
School of Management
3399 North Road
Poughkeepsie, NY 12601
ak2912@columbia.edu

1. *Introduction*

In this paper we provide an interpretive survey of recent work on endogenous information structures, where the information that is obtained by agents – who knows what – is determined at least in part by the agents communicating among themselves. This work follows our companion paper ("Part 1") discussing the economics of information where direct communication is limited, and where much of the relevant information is gleaned from making inferences based on observable actions, but is self-contained.

The problem of information control where a party (perhaps more than one) can determine what information others receive is a natural one to study when information is either abundant or can be made so (through investigations or experiments). Indeed, in a world where there is more information than can be reasonably processed, the "information problem" is no longer about lack of information, but about deciding what information to gather and to attend to. Thus, in this work we focus mostly on the perils of abundance of information, especially if this information is produced strategically or is subject to manipulation. As we discuss, such perilous abundance, often marked by strategic mis and disinformation, can undermine both markets and democratic mechanisms.

The paper is divided into six sections beyond this introduction and brief concluding remarks. The first three address communication in a world of rationality; at the center are rational individuals updating priors according to Bayes' rule. We begin in section 2 by explaining the constraints imposed by the rationality hypothesis. Section 3 discusses the burgeoning literatures on communication and information design, where an agent *designs* what information to communicate to others, in contexts where such information is not verifiable ("cheap talk") and there is no punishment for lying – other than the loss of credibility; or where communication is verifiable, but probabilistic, possibly selective and incomplete). Section 4 shows that combining direct communication with the indirect mechanisms that were at the center of discussion in the companion paper generates markedly different outcomes—most strikingly, in the standard insurance model, an equilibrium always exists and entails a pooling policy bought by both high-risk and low-risk individuals.

The last two sections focus on the more sinister side of the superabundance of information, including how mis and disinformation is leading to the polarization of society, a polarization which is best understood by going beyond the standard model of rationality and how social media and

virality are worsening polarization and giving rise to a variety of social harms (section 5); and what can be done about these harms and the increase in market power that is associated with social media—with their power, based on the use of information, undermining the foundations of competition throughout the economy (section 6).

2. *Constraints on Credibility Imposed by Rationality and Models of "Cheap Talk"*

It may seem that if the information designer can control what information others obtain, she has in effect complete freedom to choose the actions of others. She simply conveys the information that induces those she wants to manipulate to act in the way she wished. But if those she is attempting to manipulate are rational, this is, in fact, not so – it's not necessarily true that "anything goes". If the agents interpret the information they receive in a Bayesian fashion, there exist significant constraints on what the information designer can "persuade" her audience of.

Consider a simple example – suppose our information designer creates an information structure that always provides the same information – regardless of the truth. For instance, an investment adviser who always recommends purchase, or an attorney who always recommends conviction. If the agent on the receiving end knows this, she will realize that this signal is effectively uninformative and disregard it. Therefore, the information designer would not provide such a signal in the first place; in fact, the constraints (known as *Bayesian plausibility*[1], Kamenica and Gentzkow, 2011) placed on the designer by a rational, Bayesian receiver are very significant, and the work in this area nearly always assumes that these constraints are respected.

In settings where instead of providing information and then relying on agents to take the optimal action given that information, the information designer directly recommends an action, a similar mechanism is in play; the recommendation cannot be arbitrary because it has to be incentive compatible for a rational agent (who knows that the information designer is in charge of the information flow, and may have her own, ulterior, motives) to obey the recommendation; Bergemann and Morris (2016) discuss the *obedience constraint*. In much of the theoretical work on endogenous information structures, the agents know the distributions from which signals are

[1] In fact, this constraint turns out to be equivalent to a fundamental property of mathematical expectation – the law of iterated expectation, which states that the expectation of the conditional (with respect to some information) expectation of a random variable equals the unconditional expectation. Rephrasing this in the language of beliefs (estimates determined by decision makers about event probabilities), the expected posterior belief has to equal the prior belief (sometimes expressed by saying that "beliefs are a martingale"), and thus, a Bayesian agent cannot be systematically (meaning, in expectation, or on average) mislead.

drawn. Yet, the Wilson critique[2] – that knowledge of the distributions is too strong of an assumption – applies here as well; *robust information design*, which we discuss at the end of the paper, has explored the possibilities of what is and is not achievable when assumptions of common knowledge of signal distributions (and related assumptions) are relaxed.

Importantly, and much more relevantly, when decision makers are not Bayesian (a topic to which we turn in the second half of this paper), these constraints are not present; the information designer then has much more freedom to affect action via information control.

2.1. *Cheap Talk*

Implicit in much of the earlier literature on asymmetric information is the assumption that simply sending unverifiable messages couldn't convey meaningful information, and in the context in which that literature developed, that was true: less able individuals would simply say they are more able. Actions speak louder than words: because some actions are more costly (or some decisions are preferable) for low ability individuals than high ability, one can make inferences about ability by observing actions or decisions. But in other contexts, messages can convey meaningful information.

In a seminal model of costless (and therefore known as "cheap talk") communication, Crawford and Sobel (1982) establish conditions under which there is some information revelation, *even if the communication is unverifiable* - and therefore "fraud" is not punishable (as in the analysis of the companion paper). The intuition is that, *provided there is not "too much" misalignment in preferences, it is in the "sender's" interest to reveal some information correctly, and in fact, in equilibrium, this information will be believed by the "receiver"*. Of course, if the sender and receiver had identical preferences, there would be no reason for the sender not to send a truthful message, and no reason for the receiver not to believe the message. But this is true even when there is not perfect coincidence of preferences. In other words – and this is the striking interpretation of the mathematical result - there can be some degree of credible information transmission even if there is no way to verify that information and where there may be some conflict of interest.

[2] Discussed more fully in the companion paper.

Important in this literature are the assumption of *commitment* (or lack thereof), and the notion of an *information disclosure strategy*. An information disclosure strategy is a mapping that reveals signals (possibly with some randomness) that depend on the state of the world. In the cheap talk literature one assumes that the sender in *not* able to commit to such a strategy before he sees the relevant information. This is one way in which cheap talk differs from information design, where the sender *can* commit to a disclosure strategy.[3] In a sense, such noncommitment is more realistic – if the receiver doesn't observe the signals, it's hard to verify whether the sender has complied with the strategy he has committed to (and in general, the sender may wish to deviate from the strategy he committed to, upon observing more information). An assumption of commitment to telling the truth (or any more general information disclosure strategy) before observing any information is questionable if, once information is observed, there is always an incentive to break the committed-to strategy.

Subsequently, this model has been used to incorporate many other elements: costs of misrepresentation ("lying", Kartik (2009))[4], multiple dimensions of information (such as would arise if there are for instance, incompatible considerations, say, climate damage and private costs, (Battaglini, 2002), multiple senders (Ambrus and Takahashi, 2008), partial rankings or comparative (as opposed to absolute) statements (Chakraborty and Harbaugh, 2007), and repeated interactions (Ambrus, Azevedo, and Kamada, 2013). Sobel (2013) provides an insightful and exhaustive overview of the literature up to that point. Typical results state conditions under which *some* information revelation is an equilibrium outcome, and study the implications of these conditions (such as what will be communicated, what kinds of assumptions and reasoning are necessary and sufficient for communication, and what the preferences of the agents must be for communication to occur). The results are sometimes of the flavor "one needs at least this much alignment in preferences" for information revelation; generally the "closer" the interests of the agents are, the closer the equilibria are to full revelation.

Crawford and Sobel (1982) also provide a simple example, and one that is widely used in applications (in particular, in experimental economics); the sender is privately informed about (i.e.,

[3] If the sender could, instead, commit to an information disclosure strategy before observing her information, the interpretation of "cheap talk" would no longer apply. Because the message is now credible (it's coming from a communication device, the parameters of which were fixed before any private information was observed), the communication now has a "hard information" flavor.

[4] The costs of misrepresentation are "internal," i.e., not imposed by a third party, as in the case of fraud laws.

observes) the unknown payoff-relevant state - simply a real number between 0 and 1, while the receiver knows only that the distribution of the number, which is uniform. The sender sends a costless, unverifiable message (that is intended to convey some information about the state, but may be used to deceive the receiver). The receiver wishes to take an action (choose a number) that is as close to the true number as possible, while the sender has an upward bias – she wants the receiver to "inflate" the number the receiver chooses. Because the distribution of the states is uniform, and the preferences are quadratic in the difference between the action and the state, this is known as the "uniform-quadratic" setting. The bias is capturing the misalignment in preferences (such as would arise say, in financial advice or sales applications). There is always an uninformative ("babbling") equilibrium, where the receiver disregards any message, and no information is conveyed. (This equilibrium exists in many, if not all, models of cheap talk.) An imperfectly informative equilibrium of this model (what Crawford and Sobel refer to as a "partition" equilibrium) takes the following form: all sender types below a certain cutoff (i.e., senders who saw a low number) send a low message, while higher types (i.e., senders who observed a higher number) choose a higher message. For instance, in this "uniform-quadratic" setting, sender types above 0 and below 12/30 all send the message "low", (and the receiver takes the action 6/30, averaging over these states) while sender types above 12/30 (and below 1) send the message "high" (and the receiver takes the action 21/30, reflecting the expected state conditional on it being in the reported interval). Thus, there is some information conveyed: the senders inform the receiver that the true state is within some interval (the element of the partition), but do not state the exact number. The receiver is thus better informed than in the babbling equilibrium (his action is closer to the true state than it would have been under no communication), but is not perfectly informed. Of course, the receiver is also never systematically (meaning, in expectation, *ex ante*) misled; as we have already noted, systematic deception is impossible to obtain in standard models. There may be many such equilibria, differing in the number of such partitions; the more partitions there are, the more information is communicated. All equilibria are essentially equivalent to a partition equilibrium, and later Chen, Kartik, and Sobel (2008) provided an equilibrium selection condition ("no incentive to separate") that selects the most informative equilibrium.[5] In the Crawford and Sobel (1982) model, if there exists an equilibrium with some

[5] The "no incentive to separate" (NITS) is a condition that equilibria of cheap talk games may or may not satisfy. An equilibrium satisfies NITS if in that equilibrium the lowest type of sender prefers her equilibrium payoff (in that

information revelation, there also exists another one with less information revealed (for example, if there exists an equilibrium with a partition of the [0,1] interval into four elements, there is also one with three elements, and one with two elements). The latter are less informative, since there are fewer actions that can be induced, and therefore, on average, the actions will be farther from those appropriate for the state.

In an important example of cheap talk with *multiple* senders, as opposed to one sender, as in Crawford and Sobel (1982), Battaglini (2002) leverages conflict of interest *across* dimensions of information to – strikingly – obtain full revelation (although Ambrus and Takahashi (2008) qualify this result by pointing out the necessity of the richness of the space across which dimensions of agreement can be exploited to obtain full revelation).

In Battaglini's (2002) leading example, there are two senders ("experts") – one with expertise in carbon emissions and another with expertise in economic policy. The receiver has to take an action that is two-dimensional: a recommended level of carbon emissions, and a tax/subsidy policy. Both senders have preferences over both dimensions; their ideal points may be arbitrarily far apart (and thus the conflict of interest with the receiver may be arbitrarily large). Battaglini (2002) constructs a fully revealing equilibrium in the following fashion: Both senders send two-dimensional signals. Suppose that they both tell the truth, and the receiver implements the first dimension of the first sender's (two-dimensional) message, and the second dimension of the second expert's message. It turns out that (unless the experts' ideal points lie on a line in two-dimensional space – a very restrictive condition that is unlikely to be satisfied in applications) neither expert has an incentive to lie, and the receiver obtains full revelation, even in this setting of unverifiable information and preference conflict. For intuition about why this is true, consider the following two-dimensional example: suppose that the true state is *(0,0)*, that the first sender is

equilibrium) to the payoff that she gets if the receiver (somehow) knew her type, and responded optimally. The intuition for this is this: in cheap talk games, because of the upward bias of the sender, every type of sender above the lowest prefers to be revealed as itself. Thus, the "worst" thing the receiver can believe upon observing some message is that it is coming from the lowest type, which would yield a low payoff for the sender – so certainly, the higher types prefer their equilibrium payoff to deviating to something else, if that will be interpreted as coming from the lowest type. (The focus on beliefs about the lowest type is due to the fact that in other signaling games, all equilibria will remain equilibria if the receiver interprets unsent messages as coming from the lowest type, and thus, this is a mild restriction.) If even the lowest type prefers the equilibrium payoff to deviating (and being treated as itself), the equilibrium is, intuitively, robust to deviations. Alternatively, if the NITS condition were not satisfied, the low type would prefer to deviate from the equilibrium, breaking it; the only way for this to happen is if the receiver's action is "far" from the lowest type's optimal point – which, in turn, can only happen in an equilibrium with coarse partitions and little communication. Under standard regularity assumptions, Chen, Kartik, and Sobel (2008) show that the only equilibrium that satisfies this condition is the one with the most communication.

biased in his preferences by one in the vertical direction only (so that his preferred point is *(0,1)*, and that the second sender is biased by one in the horizontal direction only. Suppose that in equilibrium, the receiver implements the *horizontal* (opposite of the bias dimension) coordinate of the first sender, and the vertical coordinate of the second sender. Consider now the incentives of any agent, say, the first, to lie: if the second agent will tell the truth, the vertical coordinate of the induced action will be zero. Thus, the first agent will be constrained to choosing the best (for him) point on the horizontal line through zero. Because, as we assumed the first expert is biased only in the vertical direction and that the first coordinate of the true state is zero, the best that the first expert can do is tell the truth, and report *x=0*. But then the true state, *(0,0)*, is implemented. The construction is exploiting dimensions of *dis*agreement between the senders (changing the coordinate system to fit this example, if necessary): asking each expert to report the coordinate along which the *other* expert is more biased. Of course, this example is very special; the original paper generalizes it and shows that the intuition holds much more generally.

2.2 *Endogenizing Information and Conflict of Interest in Cheap Talk Models*

It is intuitive that if the sender doesn't use all the information he has and if it is costly to obtain information, he would not obtain information he did not use, in which case, the Crawford-Sobel results would have to be modified. Argenziano, Severinov, and Squintani (2016) and Pei (2015) study cheap talk with endogenous information, where information can be obtained at a cost. The difference between their settings lies in the modelling of the sender's information: Argenziano and co-authors use a sender who chooses how many repeated Bernoulli experiments to conduct, while Pei allows the sender to choose partitions of the state space.

Argenziano and co-authors study two variations: a game that proceeds *á la* Crawford and Sobel where the receiver observes how many trials the sender has conducted and a game where the receiver does not know this. They find (under conditions on parameters of the game) inefficient overinvestment in information – the sender's precision is too high in both variations (provided the expert's bias is not too large); i.e., the sender acquires more costly information than the receiver would, if she could do so, in all Pareto-efficient equilibria (although there also are Pareto-inefficient ones). Overinvestment occurs even in a setting where the sender is unbiased – for any number of Bernoulli experiments he chooses to acquire, the optimal action of the sender and receiver coincide.

If the receiver observes the number of trials the sender has performed, the result is driven by "equilibrium pessimism" of the receiver: she adopts a skeptical posture (see also Milgrom and Roberts, 1986), completely ignoring the transmitted information (which is bad for the sender), unless the sender acquires the equilibrium amount of information (which is too high). The intuition for overinvestment in information in this case goes as follows. Key are two observations: the only way the receiver has to induce the sender to do the preferred number of trials is to punish him for not doing so; and the only punishment is not paying attention to the tests. Moreover the receiver wants the benefits of the tests, but doesn't bear the costs, so for the receiver punishment strategies that lead to overinvestment by the sender are costless. Suppose the receiver wished to induce a number of trials that is just barely greater than the efficient number (i.e., that the receiver would have performed, if he could, and had to pay for). If the receiver uses a strong enough punishment in case of a deviation to a lower number of trials, the sender has an incentive to perform the required, overly high, number of trials (because if he doesn't the punishment is strong). The deviation to the "correct" number of trials is less profitable than the on-the-equilibrium path overinvestment (note that this does not hold for numbers of trials that are *very* large relative to the efficient number; one cannot induce anything one wants). Because the costs of these trials are borne by the sender (and not by the receiver), the equilibrium number of trials is higher than it would have been had the receiver performed them: the receiver prefers a higher number of trials when he doesn't have to pay for them, resulting in overinvestment in information acquisition.

If the receiver does not observe how many trials were performed, *in equilibrium*, the receiver believes that the sender has acquired the equilibrium amount of information even if she has not. In this case, the reason the sender does not deviate to a smaller number of trials is that if he does, his information is also worse, and it is this loss in precision, combined with the fact that the receiver expects the signal to be precise, and the fact that their interests are not completely misaligned, that drives the overinvestement result. The sender is incurring a loss in utility from the fact that the sender's utility is affected by the receiver, and the receiver's actions differ from those the sender would like him to take.

In closely related work, Pei studies a cheap talk setting where the sender may become more informed (in the Blackwell sense)[6] at a cost. Consistent with the intuition above, Pei finds that the sender always communicates all of the obtained information. However, equilibrium may be more

[6] See the companion paper for a discussion of Blackwell informativeness.

or less informative than the Crawford-Sobel one. Thus in a less informative equilibrium, communicating "everything she knows" may entail communicating more or less information than what is communicated in Crawford-Sobel's model with an exogenously informed sender in the most informative equilibrium of their game.[7] The results from this theoretical literature (with direct communication) reinforce the conclusions we highlighted in the companion work ("Part 1" with indirect communication) – the presence of incomplete, endogenous, and asymmetric information (with endogeneity in both the acquisition of information and its communication) makes a great deal of difference relative to the exogenously determined, symmetric and perfect information benchmark. Information may not be acquired or communicated efficiently. Endogeneity in information acquisition may not result in the "optimal" amount of information, and may, instead, serve as yet another source of inefficiency. Both overinvestment, and underinvestment in information acquisition may occur. Importantly, whatever information is communicated, will not, in general be entirely truthful, or precise. Endogenous communication is suboptimal, often too "coarse," even in the case when both agents would be better off if the sender could (somehow) commit to tell the truth. (Recall that in the cheap talk world, commitment is assumed away.)

3. *Bayesian Persuasion and Information Design*

The communication problem that we have been discussing is now viewed as part of a broader research agenda referred to as information design. When one party has more information about a variable of general relevance to others (as in the analysis of cheap talk of the previous section), it can choose what information to release. The informed party wants to do so to induce others to act in ways that maximize its utility. The *information design* problem is defined by Taneva (2019) as follows (italics in original)[8]:

> *Mechanism design* takes the informational environment as given and focuses on providing incentives for desired equilibrium behavior by committing to an extensive form of the

[7] Antić and Persico (2020) endogenize the difference in preferences in the cheap talk model – the magnitude of the "conflict of interest". Recall that in the workhorse Crawford-Sobel model, the preferences of the sender and receiver differ by a "bias parameter", interpreted as a difference between the sender-optimal and receiver-optimal actions given any information. The work of Antić and Persico can be thought of endogenizing this bias parameter (although their setting in much more general); agents can "purchase" elements that change their utility functions. In general, information transmission will still be partial in their setting.

[8] For a literature review, see Bergemann and Morris (2019). Mechanism design is discussed in the companion paper.

strategic interaction, i.e., a mechanism. In contrast to this, information design studies the way a designer can manipulate the equilibrium behavior of agents by selecting the informational environment under which they operate while holding the mechanism fixed. *Information design* thus applies to situations where a designer is able to influence the optimal behavior of agents only through the information she provides about the state, without being able to change any aspects of the mechanism.

A typical question of mechanism design is: For some specification of payoffs, what are the outcomes as a function of information structure (a set of signals for agents that they use to update their beliefs and play the game)? What payoffs (and beliefs) are feasible? And among these, which mechanism optimizes the wellbeing of the mechanism designer? By contrast, information design focuses on the selection of information structures; as we've explained, an information structure specifies what information is transmitted as a function of what information (i.e.., which state of the world) was observed. One critical difference between the cheap talk literature and the information design literature is that in the latter, the information designed can commit herself to an information structure, while in the former they cannot.[9] Of all the possible information structures, which is optimal, i.e., maximizes a given objective function, given that agents will obtain information through this information structure and then act "selfishly"? Generally, the information designer cannot provide information that always incentivizes the agent to act in a particular way; however, it is possible in many cases to significantly influence behavior. In the typical view, the information designer *commits* to an information disclosure rule that is state dependent (before the designer herself observes the state), and possibly probabilistic, the state is realized according to a prespecified commonly known distribution, the signals are realized according to the distribution chosen by the information designer (this is also known by the agents), and agents proceed to play a game. Because the information designer committed to the information structure *before* she observed the state, the signal realizations are now credible (in the sense of being drawn from a distribution, and not subject to manipulation by the designer). But of course, if the signal distributions are nondegenerate (as is typically the case), these signals are also possibly wrong (because of the stochasticity); the agents take this into account, as does the designer when choosing the distributions. This literature elucidates the limits of what exactly is possible, and how

[9] We discuss the assumption of commitment, and reasons why it is questionable, in section 2.1

this depends on the preferences of those that the information designer wishes to influence and on the environment.

3.1 *Information Design, Commitment and Persuasion*

Kamenica and Gentzkow (2011) illustrate the possibilities and limits of information design, taking into account the fact that if agents are rational (in particular, use Bayes' rule), one cannot mislead them systematically. They observe that one can provide information so that they act in a way that is at least probabilistically much more favorable to the information designer. As noted in the earlier discussion of cheap talk (see also Bergemann and Morris (2019)), it may be optimal to release some but not all the information, i.e., to partially obfuscate. By doing so, one can "persuade" agents to take actions that are more favorable to the informed party - actions that she might not otherwise have undertaken. "Persuasion" within the information design framework is understood to mean provision of strategically designed information to influence action.

As discussed earlier, this literature typically *assumes* commitment power on the part of the party controlling the information and in this way differs from the cheap talk literature; commitment refers not just to the provision of evidence but also, for instance, the design of an investigation or a clinical trial. (The literature makes an important distinction between information control and commitment power. "Information control" typically refers to settings where the sender can provide more information than she has initially, by designing appropriate experiments, and "commitment power" refers to the fact that the sender cannot hide the signal realizations once they are realized – but the *probability* with which they are realized in each state is under the control of the sender. In other words, the sender commits to running an experiment, the results of which will be observed whether they are favourable to the sender or not.),

Typically this literature assumes common prior beliefs about the underlying payoff relevant state. Hedlund (2017) and Kosenko (2022) extend the persuasion problem to include private information on the part of the sender: The sender may have more precise (but perhaps still imperfect) information than the receiver when designing the experiments (and, crucially, the sender is the only party able to do experiments to obtain *more* information). In a word, the sender gets an additional informative signal about the state, that the receiver is not privy to. Hedlund (2017) shows that with such private information, if the sender can choose among all possible information structures, and they all have zero cost, private information in a model of persuasion is irrelevant – in all equilibria, either the private (imperfect) information is fully revealed, or even

more strikingly, the true state of the world is revealed. Kosenko (2022) qualifies this by pointing out that this argument relies on the availability (even if it is not used in equilibrium) of a very special information structure – one that reveals the true state of the world in every state. If this structure is unavailable (or, to put it differently, there is an arbitrarily small amount of noise, as would be the case in most applications), Kosenko (2022) shows that there are many equilibria (of varying, but incomplete informativeness), and the Hedlund (2017) result does not apply. Thus, in any realistic application, private information in a model of persuasion matters a great deal. Faced with this multiplicity of equilibria, Kosenko (2022) then provides an equilibrium selection procedure (a "refinement") that selects the more informative equilibria in this setting.[10]

Kolotilin *et al*. (2017) study persuasion with private information on the part of the *receiver* (as opposed to a privately informed *sender*, as in Hedlund (2017), and Kosenko (2022)). This makes a difference for three reasons. In general, without full information, the sender can't be sure about how the different types of receivers will interpret different signals, and respond to them. Secondly, the sender may try to elicit information about the type of the receiver. And thirdly, we wish to understand how the sender should design different communication strategies for different types, if the sender can identify who is of what type.

In principle, there can be many ways of communicating in this setting. For instance, the sender may ask the receiver about her type, or just choose an experiment for all types of receivers. Within the Kolotilin *et al*. (2017) framework (linear utilities, and binary actions for the receiver), they show that these two channels – private communication (where the sender asks the receiver for her type, and then a type-dependent experiment produces a signal), and public communication (where the sender chooses an experiment, aware of the possible private information of the receivers, and how it might affect interpretation of the signal realizations) – are equivalent. In a word, under the stated restrictions, public communication is equivalent to private communication. However, strong restrictions are necessary for this result: in general the equivalence fails.

[10] There are many different situations with a multiplicity of equilibria, and therefore, many such refinements. They differ in the strength of their assumptions (stronger assumptions on, say, the inferences that agents make upon observing messages that they should not normally observe typically yield stronger predictions – more equilibria fail such refinements) and the games to which they apply. In this setting the existing refinements turn out to fail to "refine away" the nuisance equilibria, i.e., the equilibria that don't seem to make much sense. The refinement proposed by Kosenko (2022) operates by asking agents to ascribe actions to types who benefit relatively more than other types; hence the name "belief-payoff monotonicity." Beliefs (upon observing an action) are monotonically increasing as the payoff of the type that is taking that action. This is a weak, but highly plausible, restriction on beliefs.

Moreover, Kolotilin *et al*. (2017) show that private information on the part of the receivers makes a great deal of difference; in particular, they establish an "anything goes" result – *any* interim (i.e., after observing the signal realization and her own type) utility for the receiver that is achieved (between complete information about the state, and no information about the state), can be achieved in equilibrium either by some experiment ("public" communication) or a persuasion mechanism ("private" communication). Because of this, there is no loss in generality to restricting attention to experiments (and not looking at more complex, type-dependent communication mechanisms. Finally, there are many equilibria, which differ in the amount of information transmitted. Of course, which receiver utility *will* be implemented depends on the preferences of the sender.

3.2. *Robust Information Design*

Most of the literature in this area relies on very strong, and implausible, assumptions concerning what information the parties have (namely, the distribution of the underlying state, the preferences of the agents, and how they form and update their beliefs upon observing signal realizations). There is an extensive literature exploring ways of weakening the informational iassumptions that go into information design, just as we noted in the companion paper in discussing analogous work on robust mechanism design.[11] For instance, Bergemann and Morris (2016) define "Bayes-correlated equilibria." Outcomes of Bayes-correlated equilibria (BCE) are Bayes-Nash equilibrium outcomes that could arise across all information structures such that there is a strict lower bound on the information that each agent has. In other words, the mechanism designer can be mistaken about some aspect of the environment, or the agents may even be behaving in an adversarial fashion[12] (as they are in Mathevet *et al*. (2020), where the agents are coordinating on an outcome that is worst for the mechanism designer). Yet Bergemann and Morris show that the predictions of the solution concept are reliable, in the sense that they are invariant to the information. However, as Bergemann and Morris (2016) illustrate, this power comes at a cost – BCE typically make weak predictions. Many allocations (in particular, many more than under

[11] Two important examples of other models that are robust to misspecification are Bohren (2016) and Bohren and Hauser (2021) which we do not discuss because they deal with sequential social ("observational") learning, which we do not cover in this review. See Banerjee (1992), Bikchandani *et al.* (1992) and the subsequent literature.

It is important to note the multiplicity of meanings to the term "robust." It simply refers to results that hold beyond the specifications that have previously been explored, e.g., concerning the information available to various parties.

[12] As bidders might be supposed to be, vis-à-vis the auction designer.

Bayes-Nash equilibrium) can be supported as a BCE.[13] Nonetheless, Bergemann and Morris (2016) do show that the set of BCE shrinks if and only if the informativeness of the information structure increases. An increase in the abundance of information restricts the set of equilibria.

Many of the results in this literature are technical, and, like those in mechanism design (apart from the work on auctions and matching), so far have found limited application. Sometimes, the central result is only to provide bounds on what can be achieved.[14]

Mathevet *et al*. (2020), as well as Carroll (2015) and Guo and Shmaya (2019) analyzing robust mechanism design, use a worst-case scenario approach to evaluate solutions. In the latter two papers, common prior beliefs and Bayesian updating is assumed to play no role at all. Perhaps the best way to see this new strand of literature is that it explores the opposite polar case to that which previously dominated the literature, where there is always perfect Bayesian updating. In terms of applications, Guo and Shmaya (2019a) illustrate robust information design by showing when a regulator should use a price cap versus using a subsidy, so that the policy works well (in particularly defined way) in more general circumstances than associated with common priors and Bayesian updating.

While the applications of this literature have, so far, been relatively limited, one potential area that information design may be relevant in the future relates to the platforms over which so many transactions occur, and which can observe so much behavior. These platforms clearly have much more information than others – an abundance of information. Because what information they choose to disseminate can have significant consequences, there is considerable value in understanding better what information they might choose to disseminate under various conditions. For instance, Kanoria and Saban (2017) study an example where platforms (such as dating or ride-sharing apps) can improve welfare by restricting what information agents have access to.

[13] Du (2018), Bergemann *et al*. (2016), and Brooks and Du (2021a, b, c) extend this literature.

[14] Taneva (2019) provides a characterization of the optimal information environment in the static case, using BCE as the solution concept. Mathevet *et al*. (2020) study a similar problem, but focus on hierarchies of beliefs, which they view as useful for studying "robustness, bounded rationality, collusion, or communication." One of the main features distinguishing their work is their attention to equilibrium selection; whereas most of the Bayesian persuasion and information design literature has assumed designer- or sender-optimal equilibria (reasoning that the party that moves first can "steer" the game into a particular equilibrium), they explicitly include the possibility of adversarial equilibria: those in which agents may collude on the equilibrium that is worst (not best) for the designer. This in itself is a form of robust information design

But as we note below, the objectives of the platforms and that of society may differ markedly, giving rise to the necessity of regulating platforms, a subject which we discuss further below.

3.4 *Is Information Revelation Necessarily Welfare Enhancing?*

Another arena in which information design is relevant concerns the disclosure of information by central banks, who typically have more information than other agents in the economy, both about what they may be thinking of doing and about the state of the economy. The policy discourse has centered around how much information should central banks disclose? How transparent should they be? Traditionally, they were very non-transparent, but there have been marked moves to increase transparency. This has given rise to some controversy.

Morris and Shin provide a model in which public information may have a detrimental effect on welfare (Morris and Shin, 2002), in a setting reminiscent of Keynesian beauty contests (Keynes, 1936). In this context, there may be (private) benefits to coordination. Agents have private information to which they pay insufficient attention as they overreact to public information which can serve to coordinate, so that public disclosures may lower welfare. (Their result parallels some noted in the companion piece, where more information may be welfare decreasing.)

However, Svennson (2006) shows that even in their restrictive model, that is not likely to be the case: under plausible assumptions about the precision of the information available to private actors versus that of the monetary authorities and the ambient noise in the economy, *greater transparency is welfare improving*.[15] More generally, whether the conclusion of Morris and Shin's 2002 work supports or opposes public release of information (i.e., transparency) depends on the model specification. Angeletos and Pavan (2004) find that public information may decrease or increase welfare, depending on the strength of complementarities, the link between individual returns on investment and aggregate return; if the complementarities are weak, more transparency in public information increases welfare. If, on the other hand, the complementarities are strong, there may be multiple equilibria, in some of which greater transparency is welfare-decreasing. Angeletos and Pavan (2007) conclude:

> Because we allow for various strategic and external effects, there is no simple answer [to the question of whether more transparency is welfare increasing]. For example, there are

[15] Morris, Shin, and Tong (2006) offer a response to Svennson (2006) in which they emphasize the importance of different criteria of efficiency for evaluating outcomes.

economies where welfare would be higher if agents were to raise their reliance on public information and economies where the converse is true. Similarly, there are economies where any information is socially valuable and economies where welfare decreases with both private and public information. This is consistent with the folk theorem that "anything goes" in a second-best world.

This is a large literature. In macroeconomics beauty contests appear in the literature on monetary policy (Woodford, 2002), and business cycles (Angeletos and La'O, 2010; Benhabib, Wang, and Wen, 2015); Zhuo and Pedroni, 2020) provide a unifying framework for this work. Jackson and Pernoud (2021) survey some of the related literature on financial fragility, while Blanchard (2009), Angeletos, Huo and Sastry (2021) and Goldstein and Liyan (2017) offer additional reviews on the different facets of the question of the value of information disclosure in macroeconomics and finance.

3.5 *Beyond Standard Information Design*

In practice, persuasion and selective information disclosure may be greatly affected by (rational) understandings of individual irrationalities.[16] There is by now a large literature in behavioral economics showing that at least in a wide variety of circumstances, individuals do *not* behave in the way assumed by the literature on information design (in particular, in forming expectations using Bayes' theorem). It is this reality that has given rise to the huge advertising industry (with much, if not most, of the "information" provided being not informative at all) and to the problems of mis and disinformation to which we turn below.

The information design literature has focused on one aspect of communication: what information at the disposal of an informed party to disclose to others. There are several related issues, some of which have been touched on in the earlier literature already referred to in this and the companion work: (a) What information should the parties themselves gather (as in Stiglitz (1984), where an initially uninformed individual in equilibrium acquires information to make himself more informed than his trading partners), and of the information they have, what should and can they credibly convey, and how (the issue of information design associated with direct communication upon which this section has focused)?; (b) What actions should the uninformed take to extract information from the informed (as in the screening literature)?; (c) What actions

[16] Akerlof and Shiller (2015) explore related ideas.

should the informed take to convey whatever information they decide convey (i.e., signaling through actions – going beyond direct communication); (d) What actions should the informed take to make it more difficult to extract such information (as in the Edlin and Stiglitz (1995) analysis where managers make portfolio decisions that make it more difficult for outsiders to assess the net worth of the firm)?; (e) What actions should the government undertake that affect communication - both explicit disclosure rules (with penalties for incomplete disclosures) and policies that affect payoffs (e.g. "mechanism design") and therefore incentives for disclosure?; and (f) How transparent should the government itself be, i.e., how much of the information at its disposal should it disclose? In short, asymmetries of information (both *ex ante*, before communication, and *ex interim*, after communication, but before all uncertainty has been resolved) are endogenous and need to be modelled. Moreover, of increasing importance are the consequences (including potential penalties) of providing mis and disinformation going beyond partial information disclosure. And the behavioral responses have to be assessed in models which do not assume full rationality. We turn to this subject shortly.

4. *Combining Communication and Indirect Inference*

In the companion paper, we noted a conundrum: if there exist secret contracts, there appeared to be no competitive equilibrium in the standard model with asymmetric information. The standard adverse selection price equilibrium could be broken by a firm offering a large quantity contract (knowing the individual who is purchasing the contract) at a price slightly below the pooling price and make a profit. And the standard Rothschild-Stiglitz (1976) exclusive contract quantity equilibrium can't be enforced, because of the secret contract, and it turns out that the separating equilibrium that they identify doesn't work: high risk individuals buy the contract intended for the low risk individual and supplement it with secret insurance. Kosenko, Stiglitz, and Yun (2023, KSY for brevity) show more generally that there never exists an equilibrium: every proposed set of separating contracts can be broken, every proposed pooling equilibrium can be broken, and there is no hybrid equilibrium

KSY go on to investigate an equilibrium in which individuals and insurance firms *can* communicate directly anything (that they know) with each other. This stands in marked contrast to the earlier work of both Akerlof (1970) and Rothschild-Stiglitz. There, individuals' only information derives from what they themselves observe. In the case of the adverse selection

equilibrium, insurance firms only observe whether the individual has bought a policy at a particular price—this conveys a limited amount of information. In Rothschild-Stiglitz, firms observe that a particular individual chose a particular policy from a known set of policies; this can convey more information, though that information too is limited.

In KSY, individuals can't lie, but they don't have to be fully forthcoming. The remarkable result is that with the seemingly small change in assumptions, allowing secret contracts and direct communication between consumers and firms, all the standard results are reversed: Under very weak conditions,[17] there *always* exists an equilibrium and it entails partial pooling, with the low- and high-risk individuals purchasing a common policy (the pooling policy that maximizes the utility of the low-risk individual) and the high-risk individual purchasing supplemental (undisclosed) insurance to become fully insured.[18]

One prominent model of thinking is the level-*k* model and the closely related cognitive hierarchies model.

In level-*k* models (Stahl and Wilson, 1994, Nagel, 1995, Costa-Gomes, Crawford, and Broseta, 2001) all agents in a game have a type (level-0, often called L0; level-1, called L1; level-2, called L2, and so on), each type occurring with a pre-specified probability. Higher types are more sophisticated in their reasoning. L0 is the least sophisticated type; they are non-strategic, or as Crawford, Costa-Gomes, and Iriberri, 2013, put it, they have "a strategically naïve initial assessment of others' likely responses to the game." They take an action specified by the modeler (often their choice is modelled by assuming that they take all actions with equal probability – a uniform random choice). L1s best respond to the L0 types. L2s best respond to L1s, L3 best respond to L2s, and so on. Thus, in level-*k* thinking models agents best respond to the level of reasoning sophistication below them.[19] (Note than this is an explicit process for how players reason when deciding – a model of *thinking*. This model has the potential to out-predict what is called *equilibrium* behaviour (i.e., behaviour in which there is *full* rationality, taking into the infinite set of inferences described above) in novel situations.[20] Empirically, most players' behaviour is well-

[17] Convexity of preferences is a sufficient condition. Not even the single crossing property (entailing the indifference of the high-risk individual and the low risk individual only cross once) has to be satisfied.

[18] The disclosure rule is also simple: the insurance firm discloses all of its sales to all firms that have not been disclosed as sellers of insurance to the individual.

[19] In the closely related *cognitive hierarchies* model (Camerer, Ho, and Chong, 2004), types best respond to a distribution of types below them, whereas in level-*k* models types assume that everyone is the next lowest type.

[20] Full rationality in standard game theory entails each party thinking through the infinite set of inferences, corresponding to both parties having $k = \infty$ level reasoning.

described by a relatively low level of *k* – from one to four (Crawford and Iriberi, 2007, Arad and Rubinstein, 2012).

Level-*k* models have some attractive features: players will never choose actions they know are bad for them, given some plausible conjecture about how other players are playing (namely, that opponents will never play strategies that are never a best response, given their beliefs) and can infer that others will not do so either.[21] Each player's choices are best responses to the actions of others, given their beliefs. (Bernheim, 1984, Pearce, 1984). These restrictions are less restrictive than those imposed by Nash equilibrium. Furthermore, as play continues (i.e., as the number of times the game is played grows), predictions of level-*k* models often converge to the Nash equilibrium behaviour (for instance, in "beauty contests").

There are also some drawbacks of level-*k* models – the assumption of non-strategic behaviour for the L0 types and the distribution of types give the modeler two degrees of freedom, and often, additional *ad hoc* assumptions are necessary. L0's actions are sometimes (Crawford and Iriberri, 2007, Penczynski, 2013) assumed to be specific to the game, and while the empirical frequency of L0 types is low (in fact, in applications their frequency is often assumed to be zero – meaning these types only exist in the minds of higher types), it does act as an "anchor" for the actions of others. By changing this anchor, it may be possible to change predictions. Furthermore, the sophistication levels may not be stable and fixed – there is some evidence that players endogenously change their level in response to incentives (Alaoui and Penta, 2013).

5 Mis and Disinformation, Social Media and the Polarization of Society

The revolution in economics brought about some fifty years ago by information economics focused on asymmetries in information. How informed parties can convey favourable information to the uninformed and how the uninformed parties can elicit information from the informed. In the companion paper, we discussed the role of "statements" that might convey information, noting the critical role played by verifiability. Similarly, in this work, we have explored how under some circumstances, statements may convey information, even without verifiability. In the earlier literature discussed, individuals might not disclose all the relevant information; they might take

[21] This is more formally put as each play has *k-level*-rationalizability, for finite *k.* A strategy is *1-rationalizable* if it is a best response to some strategy profile of the opponents, and a *k-rationalizable* strategy is a best response to a profile that is *(k-1)-rationalizable* .

actions to obfuscate information (as in Edlin and Stiglitz, 1995); they strategically decided on how much information and what information to release; but they could be prosecuted for lying, e.g., under fraud laws, "truth in advertising," libel laws, etc. In the absence of such laws and with costly verification, economic agents might have an incentive to lie. They may be able to "get away with it."

The verisimilitude of this literature to what actually occurs in markets and the policy questions countries face may seem weak. Most advertising is not about the provision of information—it is about preying on individuals' aspirations and vulnerabilities: the Marlboro (cigarette) man is emblematic—the hugely successful ad campaign did not provide information that smoking the cigarette would make one a rugged cowboy, information which in any case would be irrelevant for the majority of smokers living in urban areas. Phillip Morris might have provided relevant information, such as that its product was deliberately designed to be addictive or that a succession of Marlboro men had died of lung cancer, and that the smoker too might die of this or a number of other health risks arising from smoking, but chose not to do so. Advertisers attempt to induce people to buy their product, but typically not by the kind of "persuasion" that has been discussed earlier in this work.

Part of the real-world information revolution (as opposed the information revolution within the academic economic discipline) is the growth of social media, which has enhanced not just the ability to target better such advertising on those most susceptible, but also the ability to rapidly spread mis and disinformation.

5.1 *The Rationality Conundrum*

But, apart from a very limited literature within economics on fraud, little attention has been paid to concerns about mis and disinformation. This is perhaps not surprising, given economists' predilection for rationality and rational expectations. Indeed, the success of mis and disinformation represents a puzzle for standard economics, which begins by assuming individual rationality, including an individual's ability to rationally evaluate the accuracy of information and update priors, using Bayes' theorem.[22] In this perspective, individuals should put little weight on

[22] In this approach, decisionmakers are thought to start with an *a priori* probability distribution (the "prior") over the unobservable state of the world, and upon observing any information, use Bayes' theorem to compute "posterior" (or *a posteriori*) probability distributions. While this approach has the advantage of being well-formulated and mathematically tractable, as we note below, it has the marked disadvantage that it seems counter to how individuals actually behave. Moreover, the questions of how priors are formed, where they come from, whether they are shared by the agents, and how important they are, are all crucial. This approach typically abstracts from investigating these

unverified “information,” putting greater weight on sources of information that have established a reputation for accuracy. So too, presumably, information from a source that repeatedly provided mis and disinformation would lose credibility and therefore would play no role in decision making - and so would not be a problem. These are issues, of course, that are at the center of the information design problem discussed in sections 2 and 3. In particular, section 2 discussed the strong constraints imposed by the hypothesis of rationality and Bayesian reasoning.

But in fact, a central problem confronting society today is the provision and spread of mis- and disinformation, and its rapid dissemination over social media. There is a small recent literature trying to come to grips with these issues, including deriving policies that might militate against the social harms to which dis and misinformation give rise. Not surprisingly, much of this goes beyond the Bayesian framework that has been central to the analysis so far.

5.2 *Why Disinformation Matters*

This mis and disinformation is associated with high levels of social harms, including inducing people not to get vaccinated, inciting violence, and stimulating racial bigotry. The magnitude of these problems seems hard to reconcile with any model of individual rationality.

Mis and disinformation have also contribution to the polarization of society. In sections 6.5, 6.6, and 6.7, we’ll describe some of the mechanisms by which this occurs, both with and without the assumption of rationality. This polarization is rightly viewed as one of the fundamental problems facing society today.

If individuals differed only in the judgments about whether red or green lettuce were healthier, such differences would be of limited significance: those believing the red lettuce was healthier could consume more red lettuce. But there are a host of important decisions that are made collectively (including the rules that underlie any economy), and differences in worldviews are associated with major differences in views about these decisions.

The pandemic brought the issues to the fore: Requirements over vaccines and masks in the pandemic were framed by the Right as an infringement on individual liberty while the Left emphasized the importance of public health externalities, and saw mandates such as those associated with masking as appropriate regulatory responses. In the presence of externalities, there

---

questions, and starts by just assuming that there is a given prior. The following discussion suggests that priors themselves may not be as “rational”, or shared ("common”), as much of this literature assumes.

is a need for collective action. But it is hard to take the appropriate collective actions when there is the level of polarization in worldviews that is evident in many societies.

One has to ask, how can there be such differences in worldviews, when the evidence is there for all to see?

5.3 *"In a Free Marketplace of Ideas, Only the Best Win Out:" A Misguided Metaphor*[23]

There are some who say not to worry. Just as in competitive markets the best producers - the most efficient, those who produce the goods that consumers want - survive, so too in the competitive marketplace of ideas. The Greenwald-Stiglitz Theorem (1986) provides the obvious caveat: In the presence of imperfect information, markets are not in general efficient; and of necessity, the marketplace of ideas is one in which *a priori* there cannot be perfect information. The discussion in the companion paper highlighted the importance of *regulations* such as "truth in advertising," "truth in lending," and fraud laws. There is a consensus that cigarette advertising, the objective of which is to induce individuals to engage in a harmful activity, should be highly regulated, with most countries requiring some disclosure of some of the harmful effects. (One reason is that there is no adequate remedy through tort law of the harm that may follow from such advertising; even if truth eventually "wins out," i.e., even if eventually the harmful effects of cigarettes became known, those who died as a result have no adequate recourse.)

The analogy to the competitive marketplace of goods is flawed in several other ways. Typically, as we have noted, competition is limited, and that is especially so in the social media platforms, marked by high levels of network externalities. Market power, the importance of which we have already noted in this context, means lack of equal access. The intermediaries control access and money matters both in the control of the intermediaries and in getting access. Those with enough money can flood the intermediaries, including by using bots.

Moreover, the first principle of a competitive "free" market is transparency. But a market in which no one knows what messages have been sent to whom is intrinsically non-transparent. To put it another way, good information is necessary to make the marketplace for goods work. But as we explained in the companion paper, markets simply won't ensure this on their own. For instance, we regulate securities markets to ensure equal access to information in the form of the

[23] This section contains ideas from and is partially borrowed from Schiffrin and Stiglitz (2020)

SEC's fair disclosure requirement and to ensure greater access to information through a variety of disclosure requirements.

There is at least one more ingredient necessary to make markets work well: the absence of the use of force and intimidation. But unregulated trolling on social media has become a fact of life.

Thus, Schiffrin and Stiglitz (2020) concluded their discussion of the idea of a free marketplace of ideas with:

> In short, without full transparency, without a mechanism for holding participants to account, without equal ability to transmit and receive information, and with unrelenting intimidation, there is no free marketplace of ideas. One of the major insights of modern economics is that private and social incentives are often not well-aligned. If those who want to spread misinformation are willing to pay more than those who want to counter it, and if lack of transparency is more profitable than transparency, then [if we simply say] "so be it" we won't get a well-functioning marketplace of ideas […]

5.4 *Evolution, Selection, and Divergence*

There is another strand of thought that says, not to worry, but for a different reason: Those who are *more* rational will win out, they will prosper and dominate, so that eventually, through an evolutionary process, the economy converges to one well-described by full rationality, with decisions being made that incorporate all the relevant information.

As we have already noted, however, even if that were true there is ample evidence that today we're far from such a world.

Moreover, there are no strong results suggesting that the economy will converge to such a world. Indeed, even in the simpler context of competitive models, there is no assurance of convergence (Bray, 1978, 1981). More recently, Dosi *et al.* (2020) has shown in the context of a simple macroeconomic model with endogenous technological change that if market participants switch to and from simple rules for expectation formation (say simple extrapolative rules) to more sophisticated rules (least square regressions based on past data) based on their relative performance, there is not convergence to the more sophisticated rules; and that overall economic performance (both growth and volatility) is poorer with more sophisticated rules. There is no natural selection towards more rationality; and it may not even be rational to be (seemingly) more rational.

5.5. *Polarization and Mis and Disinformation in a Rational Framework: Bayes' Theorem and its Critics*

In this and the next subsection, we discuss briefly research attempting to help us understand this polarization and its persistence. We do so through the lenses both of the standard model of rationality with disparate priors and through that of modern behavioral economics. In this subsection, we provide a selective discussion of the economic literature on how it is possible that rational individuals would differ so markedly in their beliefs.

There are many taxonomies of this literature. One useful for our purpose is to divide the work according to the kind of probabilistic reasoning it uses. Many (perhaps most) models use Bayesian updating, dating back to the work of Jerzy Neyman, L.J. Savage, and Harold Jeffreys in the 1950s (and of course, of Thomas Bayes in the eighteenth century).[24]

In the companion paper, we discussed Aumann's (1976) clarification that agents can only "disagree" – have different posteriors, and know each other's posteriors – to the extent that they have different priors. A recent literature on polarization and disagreement explores the effect of assuming different priors. Sethi and Yildiz (2012) for example, assume different priors and incomplete information, and endow their agents with the ability to communicate their beliefs; they show that in this setting communication need not result in information revelation, and identify the cases in which this communication breakdown can occur. Intriguingly, they show that even if priors are heterogeneous and unobserved but correlated (as would be in a society that is in some sense relatively homogeneous), communication results in an outcome identical to one where priors are observed.

More generally, however, outcomes are less salutary. Kartik *et al*. (2021), for instance, study agents with different priors who are otherwise fully Bayesian; they show that if agents' priors are different, observing the same event leads them both to update their belief about the *other* agent's beliefs to be closer to their *own* prior – a result they dub "information validates the prior."[25]

[24] Footnote 22 raised several critical issues concerning the standard usage of priors.

[25] More precisely, they show that a more informative experiment will bring one agent's expectation of the other agent's posterior belief – upon observing a signal from the experiment – closer to the first agent's prior belief, relative to a less informative experiment. Thus, Kartik *et al* 's results, set in the context of rational Bayesian agents, are in line with those obtained in the behavioral economics/psychology literature based on confirmatory bias, where individuals discount information that is not aligned with their priors. Hoff and Stiglitz (2010) show that with confirmatory biases, there can be equilibrium fictions, and that in equilibrium, those with different priors can, *in equilibrium*, sustain their differences in beliefs. We discuss non-Bayesian belief formation more extensively below. Lipnowski and Mathevet (2018) study information disclosure to an agent with psychological biases and belief-based utility, and show how to take these into account in particular cases. De Clippel and Zhang (2020) study more general persuasion of an agent

An implication of this statistical result is that that if we start polarized, we expect more information to confirm our priors, and remain more entrenched in our own worldview.

There is still another reason for the perpetuation and amplification of differences in priors. If individuals differ in their priors, they will differ in their judgments of the accuracy of information provided by different suppliers of information. Given the scarcity of time, even if information were free they would turn to suppliers of information that are, from their perspective, "better."

Indeed, Sethi and Yildiz (2016) study a setting with heterogeneous priors and consider the tradeoff between attending to information sources that are well understood (i.e., perhaps biased, but whose bias is known) and well informed (in the sense of precision of their information). Broadly speaking, their main result is that nearly anything can happen in this setting. Many kinds of behavior, including opinion leadership (where weight is given to the views of particular individuals or sources) and information segregation "groups" (where different individuals live within different information bubbles: "individuals observe only those within their own subgroup") can arise.

Changes in technology and policy affect the extent of such fragmentation. In the era after World War II, when TV was a major media for providing new information, there were only three major national networks in the US and all aimed to provide broad and unbiased information. News programs were treated as a public service by networks (a practice that was changed, in part by the program *60 Minutes* on CBS, which showed that news programs can also bring in revenue). Fairness doctrines ensured that major different views were given airtime. Those across the political spectrum were at least exposed to similar information. But the elimination of fairness obligations in 1987 by the US Federal Communications Commission combined with cable TV and then the internet meant that the information to which those of different beliefs were exposed became markedly different. The consequences of these changes have been discussed and modelled, among others, by Glaeser (2005), Pickard (2015), Guriev and Treisman (2020), Ash, Mukand, and Rodrik (2021), and Szeidl and Szucs (2022).

The Bayesian approach has many advantages: It is mathematically tractable, and is consistent with vast amounts of research in other fields. But it suffers from a critical defect. It's long been known to be inconsistent with a wealth of evidence of behavior in a wide variety of

with non-Bayesian updating rules. They show (among other results) the optimal policy if agents distort their beliefs in a subset of non-Bayesian ways.

circumstances - humans don't typically reason about probabilities in a way that is predicted by the Bayes rule (Keynes, 1921, Allais, 1953 (presenting the famous Allais Paradox[26]), Ellsberg, 1961 (presenting the famous Ellsberg paradox), Kahneman and Tversky, 1979; for an influential early review, see Machina, 1987).[27]

5.6. *Behavioral Economics: Beyond Bayes*

None of the approaches based on Bayesian rationality, as sophisticated as they may be, can really account for the observed divergences in views. In the standard economists' model, non-scientific (for example, anti-vaccine) information would simply have no impact. The evidence is that it does, and that this is so, is consistent with a large literature in behavioral economics stressing an individual's cognitive limitations, particularly in processing statistical information and especially when such information (data) is contrary to prior beliefs.

One then has to model human probabilistic information processing in some other way; other updating procedures (for instance, probability weighting (Tversky and Kahneman, 1992, and Prelec, 1998), over/under reaction to new information (Epstein *et al*., 2008, 2010), the peak-end rule according to which intense experiences ("peak") and experiences which come last ("end") are remembered (Fredrickson and Kahneman, 1993, and Kahneman *et al.*, 1993), and attaching disproportionate weight to initial observations (Rabin and Schrag, 1999)) may be perhaps more plausible in some contexts, but are typically imposed in a somewhat *ad hoc* fashion. Ortoleva (2012) provides a prominent example of this literature in which he posits a threshold probability above which a decisionmaker acts in a Bayesian fashion, and below which (i.e., for low probabilities – "unexpected news") the decisionmaker acts differently.[28]

Twenty-first century behavioral economics has, in addition, emphasized the importance of the social formation of beliefs. Beliefs are interdependent, with information about a particular subject (such as the safety and efficacy of vaccines or the role of masks during a pandemic) interpreted through a cultural lens, which "prejudice" the assessments. Moreover, beliefs are affected by the beliefs of those with whom one interacts. This is especially so if those providing the information succeed in framing the information in ways that embed it into a cultural context.[29]

[26] Maurice Allais was the 1988 recipient of the Nobel Memorial Prize in economics.
[27] More recent literature includes De Bondt and Thaler (1985, 1987, 1990), Camerer (1998), Angrisani *et al.* (2017), Levy and Razin (2017), and Bohren and Hauser (2021).
[28] Ortoleva (2012) also cites much of the evidence for non-Bayesian behavior.
[29] See, e.g., Hoff and Stiglitz (2016), and Demeritt, Hoff, and Stiglitz (2023).

Given the polarization of views around central themes of individual liberties and collective action and the centrality of these issues to today's critical policy debates, it is not surprising that there is deep polarization around what would seem to be scientific issues like climate change or the efficacy and safety of vaccines.

While there is, at this juncture, no consensus on precisely how individuals *actually* process information to form beliefs (in contrast to the consensus over how "rational" individuals *should* form beliefs through Bayesian statistics), there is widespread understanding of some of the constitutive elements and key properties, e.g., on the importance of framing and of confirmatory bias. As Hoff and Stiglitz (2010) have shown, confirmatory bias can easily give rise to "equilibrium fictions," beliefs that are self-sustaining, even in the presence of evidence against them. And this may be even more so if individuals not only start with different priors, as in Kartik *et al* (2021) and Sethi and Yildiz (2012), and are influenced by the beliefs of others with whom they interact, but interact only or mostly with people whose priors are close to theirs but different from others.[30]

Of course, those in marketing have long sought to understand how to influence individuals' beliefs, even with "non-informative" advertising (the Marlboro Man being the quintessential example), with a modicum of success, enough evidently to justify the billions of dollars spent every year on such advertising.

5.7 Competition in Worldviews and Signal Jamming

Building on these insights, there are two important directions to be addressed in future research. One is "competition in worldviews": there is a need for a theory of metanarratives, the "lens" through which we see the world, which result in competing interpretations of the same pieces of information. How do we explain the persistence of differences in priors, even when so much of the evidence on which judgments are made is widely available?

As we noted before, the question of how competition in worldviews plays out is of crucial importance. The fact of the matter is that many critical events happen with such rarity that there is ample opportunity for differing interpretation of their origins and consequences (Guzman and Stiglitz 2020, 2021). The analysis of the preceding subsection helps our understanding of how differences in worldviews could persist.

[30] A key aspect of the work cited in fn. 25 is the social formation of beliefs.

Szeidl and Szucs (2022) present a political economy model where, although information is verifiable (there is an "objective reality" in their world), some politicians are nevertheless able to persuade some voters of something false. The key assumption of their model is that some voters may believe a verifiably false message, because they believe that with some small probability that message could be true.[31] In other words, "propaganda makes the voter assign positive probability to a nonexistent alternative reality" (Szeidl and Szucz, 2022). One of the key theoretical findings of their framework is that once such an alternative reality is created, it can persist, even in presence of clear evidence of its falsity, and will cause the persuaded voter to act against her best interest.

A second promising line of thought is "signal jamming" as information manipulation. A malicious actor may exploit the fact that audiences have a limited capacity for information processing. The possibility of such obfuscation has been critical in the design of disclosure requirements. (See the discussion of the companion paper.) Burying the required disclosure (nutritional information for food, risk for investments) in a barrage of other information attenuates the value of the information being provided. The audience may then either tune out informative signals, or, once information processing capacity has been reached, tune out all signals. In other words, one may "crowd out" informative signals by providing a surfeit of uninformative ones, with the aim of limiting the ability to process the informative ones. In some cases, e.g., in the health hazards of cigarettes, governments have specified how the relevant information is to be disclosed to prevent such obfuscation.

6 *Social Media, Social Divisions, and Public Policy*

Social media platforms have been able to take advantage of not only advances in AI, but of understandings of human behavior and information processing in ways that have increased their profits while imposing large costs on society, including exacerbating societal polarization. They have a business model that profits from engagement (sometimes through enragement) and. AI algorithms have targeted different individuals with different information designed to enhance engagement, fragmenting the information structure beyond anything that had previously been

[31] This is a very different use of the word "persuasion" than that employed earlier in section 3. If the cost of verification were really zero, they would engage in verification (even if they thought there was some probability that the information was true), in which case the false message would have no impact. But, in practice, there are costs, e.g. of attention diverted from elsewhere.

possible and in ways that have enhanced polarization. The ability to create separate communities, reinforcing the disparate beliefs, has made matters still worse.

Virality meant information could spread quickly, more quickly than "antidotes" to the misinformation could be designed. The lack of transparency in who gets what messages has meant that the antidotes could not be effectively delivered in the relevant time span, if at all. In addition, the lack of transparency of the algorithms and of the training data used by AI, and especially by the large language models (which may eventually contribute to generative artificial intelligence, a kind of AI that can create many kinds of data and media) have created a possibility where even completely incorrect information may be created and disseminated essentially without human intervention. To make matters worse, the lack of information on how such algorithms function, and information on their evolution, limits our understanding of the entire scope of the information landscape.

In the companion paper we emphasized the interaction between information and market power. Information economics helps explain the limited competition in media, including social media. Later in this section, we will explore in greater detail how this plays out with vengeance in the context of social media.

Market power, in turn, enhances the power to exploit limitations in information. Another central theme of modern economics is the link between economic power and political power - a vicious circle where the concentration of economic power leads to a concentration of political power, which results in rules of the game enhancing the concentration of economic power.[32]

It not only pays the very rich to create a "narrative" that supports an economic environment that enriches them (say with low taxes and the ability to engage in exploitation of others), but many of them also come to believe this narrative.

That is why market power in the media may be particularly invidious. It gives those with the wealth and the desire to control dominant media the power to shape the societal narrative (discussed in 5.7) in ways which affect the interpretation of data.[33] This effect may be even stronger than the more narrowly defined distortions in the information disseminated discussed above; it is this which enables the success of mis and disinformation campaigns. And when there is not sufficient media diversity, the ability to counter the narrative is limited. But even with some

[32] See Stiglitz (2012, 2015, 2019).
[33] See Prat (2018) for a definition of "media power" and an application to the case of the US.

media diversity, the polarization effects described earlier mean that even if there are outlets providing counter-narratives and "true" facts, their impacts may be limited.

The invidious effects of social media are even greater, not just because competition may be more limited and they have a greater ability to target information, but also because of policy. In the US, social media platforms are shielded by section 230 of the Communications Decency Act (1996) from liability for what they transmit across their platforms, in a way that standard media are not. Conventional media are subject, for instance, to being sued under libel and fraud laws; not so for the platforms. A provision originally designed to encourage a nascent industry has led to the viral dissemination of mis and disinformation, the consequences of which have been discussed extensively elsewhere.

6.1 Regulating the Societal Harms of Social Media and Mis and Disinformation in a Democratic Society

Today, many countries, recognizing the variety of societal harms arising from mis- and disinformation, especially over social media, are debating how to regulate them.[34] The EU, for instance, has adopted the Digital Services Act. A central question in the design of such regulations (a full discussion of which is beyond the scope of this paper) is how to prevent such harms within democratic frameworks that emphasize, for instance, freedom of speech (First Amendment rights). Societies including the US have not, however, taken absolutist positions: there are prohibitions against fraud ("lying" in commercial contexts where the lie results in harm), false advertising, child pornography, and crying fire in a crowded theatre. Some countries ban hate speech. Clearly, if there are greater harms emanating from mis and disinformation on, say, social media, the "balancing" entailed in the design of regulations changes, towards greater intervention, especially when such intervention is directed at the extent of virality.[35]

Indeed, in some way, there is already a form of regulation of virality, but it is effectively regulation delegated to the platforms themselves—virality is determined by the social media platforms (through their algorithms) in ways which are not transparent but which maximize their profits, almost regardless of the social harms generated, with a limited role given to content

[34] There is a nascent literature on the subject in media studies.

[35] Note that historically, censoring has sometimes taken the form of "regulating" virality, by insisting that adverse news be placed in a position in the newspaper where it is less likely to be read.

moderation. There is a growing consensus that if social media and mass online interaction results in harm, self-regulation is no more effective in this context than it was in financial markets.

6.2. *The Market Power of Social Media and the Need to Control It*

The enormous profits of the social media companies are a strong indication of the lack of competition. It is not that the successful companies are necessarily that much more productive or innovative than their rivals. (Indeed, in many cases, they made only small but still important innovations that gave them some advantage over rivals.) Normally, such large profits would attract entry, which in turn would lead to the dissipation of the profits. This, however, has not happened. The reason is simple: network externalities. The value of being on a platform like Facebook depends on the presence of others being on the platform. It is hard, in such a situation, to displace even a relatively inefficient incumbent, one who does not serve the interests of those on the platform as well as others might.

But there is another element to the vicious circle which has given rise to their enormous market power and profits. Their business model, based on the use of information garnered from interactions that occur over their platform, has been a double-edged sword. (A key element in this process is that this data creation --data that is then monetized by the platforms-- by virtue of mere interaction with the platform is essentially independent of the content of the interaction.) The use of the greater information that they have has allowed them to better target messages in ways that engender more engagement (and thus generate still more information). With attention (and time) a scarce commodity, "better" targeting could mean individuals receive messages that are more relevant, thereby leading to more efficient resource allocations (purchases that result in individuals enjoying a higher level of well-being). That, however, is not the *objective* of the better targeting. The objective is more profits, which are derived from advertising revenues, which in turn are derived from inducing more profitable purchases. Increased profits from sales, in turn, can result from more effective price discrimination - capturing more of the individual's consumer surplus. They can also arise from increased sales, including to people whose weaknesses they are taking advantage of, e.g., to gambling addicts.

In addition, a platform can increase its profits by increasing its competitive advantage over rivals by "hoarding" information, enabling it to engage in this targeting better than others. Of course, if the information were used in a socially productive way, economic efficiency would

require its sharing, since information is a public good. (See companion paper.[36]) Hoarding such information, while privately profitable, is doubly inefficient because it not only prevents its full use, but endows the platform with market power. With data being a key (largely unpriced) resource, especially important in artificial intelligence (AI), there is a vicious circle. Larger platforms get more data, which give them a competitive advantage over rivals, enhancing still further their market power, with profits often generated from their better ability to extract consumer surplus out of customers, not their better ability to serve consumers.

There are tensions, of course, between the efficient use of information, the anti-competitive hoarding of information, and privacy concerns. In the standard competitive market, there is no value to information about consumer preferences. As we discussed in the companion paper, in the standard setting, price is a "sufficient statistic" for conveying all relevant information. But while such information has no incremental private or social value in a competitive market, in the real world, with imperfect competition and incomplete markets, it can be enormously valuable to a firm, increasing significantly its profits.

While it is difficult to ascertain the extent, if any, of improvement in resource allocation resulting from the information being exploited by the platforms, one analytic result is clear: the use of this information to engage in price discrimination undermines the standard argument for the efficiency of competitive markets (the first welfare theorem), which is premised on every household and firm facing the same prices.[37]

There is another adverse by-product of the business model of social media which entails maximizing engagement: engagement is enhanced by enragement, and especially of a kind that has been associated with polarization (discussed above). Thus, one of the societal harms of social media, as it operates today, is that it has created a more divided society, making cooperative actions

[36] Because information is a public good, disinformation, a public bad, it may not pay anyone individually to stop its production and dissemination or to engage in activities (like showing its untruthfulness) that might undermine it. Stopping mis and disinformation is a public good. Without public action, there will be an undersupply of efforts at countering mis- and disinformation. There is thus a strong argument for fraud laws, even if such laws might be viewed as an infringement on "free speech," interpreted in an absolutist way. Even more so, there is a strong argument for laws restricting virality.

[37] Although efficiency can also be sustained by *perfect* price discrimination, the information generated by platforms and employed by firms engaged in price discrimination is far from sufficient to enable perfect price discrimination (although modern algorithms can come much closer than before). Stiglitz (1977) showed that, in the presence of imperfect information, the welfare losses associated with monopoly arise from the attempt to engage in imperfect price discrimination.

to address society's common problems far more difficult. There is an obvious social externality—but one which the social media companies, in their search for profits, pay no attention to.

Importantly, the informational advantages that are obtained by the large technological firms of today are qualitatively different from the monopoly or collusive advantages that the main antitrust laws enforced by the Federal Trade Commission and the Department of Justice (the Sherman Act and the Clayton Act) were designed to regulate. This is especially relevant for settings in which differential access to information confers an advantage. Because information itself is so special (one cannot "unknow" information, but one can resell a good), and because of the complex interactions between consumer and user-provided information and industry-obtained or metadata information, designing legislation ensuring competition in such circumstances is a formidable task, one the European Union is beginning to undertake in its Digital Marketing Act (DMA), the Digital Services Act (DSA), and the General Data Protection Regulation (GDPR). Such legislation, designed to promote the functioning of competitive markets with endogenous asymmetric information while ensuring appropriate levels of consumer privacy and control of user data and other basic rights (as described at the end of the previous sub-section), is necessary if the extremes of either market breakdown *à-la* Akerlof or market power are to be avoided.

7 *Concluding Comments*

The development of information economics a half century ago unleashed a revolution in economics that touched every aspect of economics. Key presumptions - that markets were efficient and that demand equaled supply in equilibrium - were undermined. The new theories provided new insights into why markets might be absent, into why governance issues were so important, and into why prices did not convey all the information that was relevant for decisions of firms and households. This, in turn, opened up new avenues of enquiry, including into the design of other mechanisms besides competitive markets for resource allocations and for the transmission of relevant information. While there has been considerable excitement about such mechanisms, their domain of applications has remained limited. These limitations, combined with the complexity of the information and mechanism design literature based on rational behavior and the limitations of even the theoretical results obtained so far, may have reinforced the conviction that what is needed is more analyses based on behavioral economics, recognizing cognitive limitations and the importance of the social determination of beliefs.

Firms in maximizing their profits, engage not only in costly efforts to overcome information asymmetries but in the selective disclosure of information, the creation of information asymmetries, and efforts to undermine the ability to overcome such asymmetries. Because what is privately profitable may not be socially desirable, governments are engaged in designing rules and regulations, including policies that affect the collection, use, and dissemination of information, to enhance societal welfare, taking into account how private actors will respond, including with respect to actions related to information. Governments (like firms) know, too, that consumers may have cognitive limitations and may not be fully rational, and they know that firms know that, and are willing and able to take advantage of these limitations. And because almost all actions (or "non-actions") can potentially convey information, public policies need to be all-encompassing.

All the problems posed by imperfect and asymmetric information have become worse with digitalization and AI. For instance, there are also pervasive *novel* externalities associated with algorithms – they may accurately predict behavior (using information on behavior of other similar users) even if this particular user has not interacted with this algorithm before. Thus, users interacting with an algorithm exert an externality on future users. This "forward" externality has not been adequately studied, nor regulated.

An essential part of the distributive battle is the battle over information; and while with perfect information outcomes in (perfectly) competitive markets are Pareto efficient, there is no such presumption that that is the case for the outcomes in these information battles. To the contrary, there is a presumption that this is not so and that appropriate intervention *can* be welfare-increasing; whether it will be depends on political processes.

What was supposed to be the information age has become the "dis and misinformation" age, with *sustained* dis and misinformation - often seemingly inconsistent with the economists' standard model of rationality. This surfeit of mis and disinformation has had marked effects on economics, politics, and society. Abundance of information has its perils. Understanding better the social interactions and cognitive functions that make such dis and misinformation so salient, and devising better policies to deal with it, should be one of the main objectives of information economics going forward.

References

Akerlof, G. A. (1970). The Market for "Lemons": Quality Uncertainty and the Market Mechanism. *The Quarterly Journal of Economics*, *84*(3), 488–500. https://doi.org/10.2307/1879431

Akerlof, G. A., & Shiller, R. J. (2015). Phishing for Phools: The Economics of Manipulation and Deception. In *Phishing for Phools*. Princeton University Press. https://doi.org/10.1515/9781400873265

Allais, M. (1953). Le Comportement de l'Homme Rationnel devant le Risque: Critique des Postulats et Axiomes de l'Ecole Americaine. *Econometrica*, *21*(4), 503–546. https://doi.org/10.2307/1907921https://doi.org/10.2307/1907921

Ambrus, A., & Takahashi, S. (2008). Multi-Sender Cheap Talk with Restricted State Spaces. *Theoretical Economics*. https://dash.harvard.edu/handle/1/3200263

Ambrus, A., Azevedo, E. M., & Kamada, Y. (2013). Hierarchical cheap talk. *Theoretical Economics*, *8*(1), 233–261. https://doi.org/10.3982/TE1038

Angeletos, G.-M., & Pavan, A. (2007). Efficient Use of Information and Social Value of Information. *Econometrica* 75 (4): 1103–42.

Angeletos, George-Marios, and Jennifer La'O. (2013). Sentiments. *Econometrica* 81 (2): 739–79.

Angeletos, George-Marios, Zhen Huo, and Karthik A. Sastry. (2021). Imperfect Macroeconomic Expectations: Evidence and Theory. *NBER Macroeconomics Annual* 35, 1-86

Angeletos, George-Marios and Pavan, Alessandro. "Transparency of Information and Coordination in Economies with Investment Complementarities." *American Economic Review*, 2004, 94(2), pp. 91-98.

Angrisani, M., Guarino, A., Jehiel, P., & Kitagawa, T. (2021). Information Redundancy Neglect versus Overconfidence: A Social Learning Experiment. *American Economic Journal: Microeconomics*, *13*(3), 163–197. https://doi.org/10.1257/mic.20180394

Antić, N. and Persico, N. (2020), Cheap Talk With Endogenous Conflict of Interest. *Econometrica*, 88: 2663-2695. https://doi.org/10.3982/ECTA14953

Ash, Elliott, Sharun Mukand, and Dani Rodrik, "Economic Interests, Worldviews, and Identities: Theory and Evidence on Ideational Politics," Working Paper 29474, National Bureau of Economic Research November 2021.

Argenziano, Rossella, Sergei Severinov, and Francesco Squintani. 2016. "Strategic Information Acquisition and Transmission." *American Economic Journal: Microeconomics*, 8 (3): 119-55.DOI: 10.1257/mic.20140135

Aumann, R. J. (1976). Agreeing to Disagree. *The Annals of Statistics*, *4*(6), 1236–1239. https://doi.org/10.1214/aos/1176343654

Banerjee, Abhijit V. (1992). A Simple Model of Herd Behavior, *The Quarterly Journal of Economics*, Volume 107, Issue 3, Pages 797–817, https://doi.org/10.2307/2118364

Battaglini, M. (2002). Multiple Referrals and Multidimensional Cheap Talk. *Econometrica*, *70*(4), 1379–1401. https://doi.org/10.1111/1468-0262.00336

Benhabib, Jess, Pengfei Wang, and Yi Wen. 2015. "Sentiments and Aggregate Demand Fluctuations." *Econometrica* 83 (2): 549–85.

Bergemann, D., & Morris, S. (2016). Bayes correlated equilibrium and the comparison of information structures in games. *Theoretical Economics*, *11*(2), 487–522. https://doi.org/10.3982/TE1808

Bergemann, D., & Morris, S. (2019). Information Design: A Unified Perspective. *Journal of Economic Literature*, *57*(1), 44–95. https://doi.org/10.1257/jel.20181489

Bikhchandani, S., Hirshleifer, D., & Welch, I. (1992). A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades. *Journal of Political Economy*, *100*(5), 992–1026.

Blanchard, Olivier. (2009). The State of Macro. *Annual Review of Economics* 1:1, 209-228

Bohren, Aislinn. (2016). Informational herding with model misspecification. *Journal of Economic Theory*, 163, issue C, p. 222-247

Bohren, J.A. and Hauser, D.N. (2021). Learning With Heterogeneous Misspecified Models: Characterization and Robustness. *Econometrica*, 89: 3025-3077. https://doi.org/10.3982/ECTA15318

Bray, M. (1981). Futures Trading, Rational Expectations, and the Efficient Markets Hypothesis. *Econometrica*, *49*(3), 575–596. https://doi.org/10.2307/1911513

Bray, M. M. (1978). *Information in Futures Markets*. Oxford University.

Brooks, B., & Du, S. (2021a). *Maxmin Auction Design with Known Expected Values*. Working paper.

Brooks, B., & Du, S. (2021b). Optimal Auction Design with Common Values: An Informationally Robust Approach. *Econometrica*, *89*(3), 1313–1360. https://doi.org/10.3982/ECTA16297

Brooks, B., & Du, S. (2021c). *On the Structure of Informationally Robust Optimal Auctions*. Working paper.

Camerer, C. Bounded Rationality in Individual Decision Making. *Experimental Economics* **1,** 163–183 (1998). https://doi.org/10.1023/A:1009944326196

Carroll, Gabriel (2015) "Robustness and Linear Contracts," *American Economic Review*, 105 (2), 536–63.

Chakraborty, A., & Harbaugh, R. (2007). Comparative cheap talk. *Journal of Economic Theory*, *132*(1), 70–94. https://doi.org/10.1016/j.jet.2005.07.003

Crawford, V. P., & Sobel, J. (1982). Strategic Information Transmission. *Econometrica*, *50*(6), 1431–1451. https://doi.org/10.2307/1913390

De Bondt, W. F. M., & Thaler, R. (1985). Does the Stock Market Overreact? *The Journal of Finance*, *40*(3), 793–805. https://doi.org/10.1111/j.1540-6261.1985.tb05004.x

De Bondt, W. F. M., & Thaler, R. H. (1987). Further Evidence On Investor Overreaction and Stock Market Seasonality. *The Journal of Finance*, *42*(3), 557–581. https://doi.org/10.1111/j.1540-6261.1987.tb04569.x

De Bondt, W. F. M., & Thaler, R. H. (1990). Do Security Analysts Overreact? *American Economic Review*, *80*(2), 52–57.

de Clippel, G., & Zhang, X. (2020). *Non-Bayesian Persuasion*. Working paper.

Dean, Mark and Nathaniel Neligh. (2017). Experimental Tests of Rational Inattention. Working Paper.

Demeritt, A. Hoff, K. and Stiglitz, J. E. (2023). *The other invisible hand: The power of culture to promote or stymie progress.* New York: Columbia University Press.

Dewan, Ambuj, and Nathaniel Neligh. (2020). Estimating Information Cost Functions in Models of Rational Inattention. *Journal of Economic Theory* 187, 1-32.

Dosi, G., Napoletano, M., Roventini, A., Stiglitz, J. E., & Treibich, T. (2020). Rational Heuristics? Expectations and Behaviors in Evolving Economies with Heterogeneous Interacting Agents. *Economic Inquiry*, *58*(3), 1487–1516. https://doi.org/10.1111/ecin.12897

Du, S. (2018). Robust Mechanisms Under Common Valuation. *Econometrica*, *86*(5), 1569–1588. https://doi.org/10.3982/ECTA14993

Edlin, A. S., & Stiglitz, J. E. (1995). Discouraging Rivals: Managerial Rent-Seeking and Economic Inefficiencies. *American Economic Review*, *85*(5), 1301–1312.

Ellsberg, D. (1961). Risk, Ambiguity, and the Savage Axioms. *The Quarterly Journal of Economics*, *75*(4), 643–669. https://doi.org/10.2307/1884324

Epstein, Larry G., Jawwad Noor, and Alvaro Sandroni. (2008). "Non-Bayesian Updating: A Theoretical Framework." *Theoretical Economics* 3 (2): 193–229.

Epstein, Larry G., Jawwad Noor, and Alvaro Sandroni. (2010). "Non-Bayesian Learning." *B. E. Journal of Theoretical Economics*: Advances in Theoretical Economics 10 (1).

Fredrickson, B. L., & Kahneman, D. (1993). Duration neglect in retrospective evaluations of affective episodes. *Journal of Personality and Social Psychology*, *65*(1), 45–55. https://doi.org/10.1037/0022-3514.65.1.45

Glaeser, Edward L., "The Political Economy of Hatred," *Quarterly Journal of Economics*, 02 2005, 120 (1), 45–86.

Goldstein, Itay and Yang, Liyan. (2017) Information Disclosure in Financial Markets. *Annual Review of Financial Economics*, Vol. 9, pp. 101-125.

Greenwald, B. C., & Stiglitz, J. E. (1986). Externalities in Economies with Imperfect Information and Incomplete Markets. *Quarterly Journal of Economics*, *101*(2), 229–264. https://doi.org/10.2307/1891114

Guriev, Sergei and Daniel Treisman. (2020). A theory of informational autocracy. *Journal of Public Economics*, 186, 104-158.

Guo, Y. and Shmaya, E. (2019a). Robust Monopoly Regulation. Working Paper.

Guo, Y. and Shmaya, E. (2019b). The Interval Structure of Optimal Disclosure. *Econometrica*, 87: 653-675. https://doi.org/10.3982/ECTA15668

Guzman, M., & Stiglitz, J. E. (2020). Towards a dynamic disequilibrium theory with randomness. *Oxford Review of Economic Policy*, *36*(3), 621–674. https://doi.org/10.1093/oxrep/graa042

Guzman, M., & Stiglitz, J. E. (2021). Pseudo-wealth and Consumption Fluctuations. *The Economic Journal*, *131*(633), 372–391. https://doi.org/10.1093/ej/ueaa102

Hedlund, Jonas. (2017). Bayesian Persuasion by a Privately Informed Sender. *Journal of Economic Theory* 167, 229-268.

Hoff, K., & Stiglitz, J. E. (2010). Equilibrium Fictions: A Cognitive Approach to Societal Rigidity. *American Economic Review*, *100*(2), 141–146. https://doi.org/10.1257/aer.100.2.141

Hoff, K., & Stiglitz, J. E. (2016). Striving for balance in economics: Towards a theory of the social determination of behavior. *Journal of Economic Behavior & Organization*, *126*, 25–57. https://doi.org/10.1016/j.jebo.2016.01.005

Hoff, K., Demeritt, A., & Stiglitz, J. E. (2023). *The Invisible Hand of Culture: How It Affects Economic and Social Progress and Can Make a Society Rigid*. Columbia University Press.

Huo, Zhen, and Marcelo Pedroni. (2020). A Single-Judge Solution to Beauty Contests. *American Economic Review*, 110 (2): 526-68.

Jackson, Matthew O. and Agathe Pernoud. (2021). Systemic Risk in Financial Networks: A Survey. *Annual Review of Economics* 13(1), 171-202

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, *47*(2), 263–291. https://doi.org/10.2307/1914185

Kahneman, D., Fredrickson, B. L., Schreiber, C. A., & Redelmeier, D. A. (1993). When More Pain Is Preferred to Less: Adding a Better End. *Psychological Science*, *4*(6), 401–405.

Kamenica, Emir, and Matthew Gentzkow. 2011. "Bayesian Persuasion." *American Economic Review*, 101 (6): 2590-2615.DOI: 10.1257/aer.101.6.2590

Kanoria, Y. and D. Saban. (2017). Facilitating the Search for Partners on Matching Platforms: Restricting the Agents' Actions. Working Paper.

Kartik, N. (2009). Strategic Communication with Lying Costs. *The Review of Economic Studies*, *76*(4), 1359–1395. https://doi.org/10.1111/j.1467-937X.2009.00559.x

Kartik, N., Lee, F. X., & Suen, W. (2021). Information Validates the Prior: A Theorem on Bayesian Updating and Applications. *American Economic Review: Insights*, *3*(2), 165–182. https://doi.org/10.1257/aeri.20200284

Keynes, J. M. (1921). *A Treatise On Probability*. Macmillan And Co., http://archive.org/details/treatiseonprobab007528mbp

Keynes, John Maynard (1936). *The General Theory of Employment, Interest and Money*. London: Macmillan.

Kolotilin, A., Mylovanov, T., Zapechelnyuk, A. and Li, M. (2017), Persuasion of a Privately Informed Receiver. *Econometrica*, 85: 1949-1964. https://doi.org/10.3982/ECTA13251

Kosenko, Andrew. (2022). Constrained Persuasion with Private Information. Forthcoming in *The B.E. Journal of Theoretical Economics*. https://doi.org/10.1515/bejte-2021-0017

Kosenko, A., Joseph E. Stiglitz, and Jungyoll Yun. (2023). Bilateral Information Disclosure in Adverse Selection Markets with Nonexclusive Competition. *Journal of Economic Organization and Behavior, 205, 144-168.*

Levy, G., & Razin, R. (2015). Correlation Neglect, Voting Behavior, and Information Aggregation. *American Economic Review*, *105*(4), 1634–1645. https://doi.org/10.1257/aer.20140134

Lipnowski, E., & Mathevet, L. (2018). Disclosure to a Psychological Audience. *American Economic Journal: Microeconomics*, *10*(4), 67–93. https://doi.org/10.1257/mic.20160247

Machina, M. J. (1987). Choice Under Uncertainty: Problems Solved and Unsolved. *Journal of Economic Perspectives*, *1*(1).

Mathevet, L., Perego, J., & Taneva, I. (2020). On Information Design in Games. *Journal of Political Economy*, *128*(4), 1370–1404. https://doi.org/10.1086/705332

Milgrom, P., & Roberts, J. (1986). Relying on the Information of Interested Parties. *The RAND Journal of Economics*, *17*(1), 18–32. https://doi.org/10.2307/2555625

Morris, Stephen, and Hyun Song Shin. 2002. "Social Value of Public Information." *American Economic Review*, 92 (5): 1521-1534.DOI: 10.1257/000282802762024610

Morris, S., Shin, H. S., & Tong, H. (2006). Social Value of Public Information: Morris and Shin (2002) Is Actually Pro-Transparency, Not Con: Reply. *The American Economic Review*, *96*(1), 453–455.

Ortoleva, P. (2012). Modeling the Change of Paradigm: Non-Bayesian Reactions to Unexpected News. *American Economic Review*, *102*(6), 2410–2436. https://doi.org/10.1257/aer.102.6.2410

Pei, Harry Di. (2015). Communication with Endogenous Information Acquisition. *Journal of Economic Theory* 160, 132-149. https://doi.org/10.1016/j.jet.2015.08.011

Pickard, V. (2015). *America's battle for media democracy: The triumph of corporate libertarianism and the future of media reform*. New York, NY: Cambridge University Press.

Prat, A. (2018). Media Power. *Journal of Political Economy* 126(4), 1747-1783. https://doi.org/10.1086/698107

Prelec, D. (1998). The Probability Weighting Function. *Econometrica*, *66*(3), 497–527. https://doi.org/10.2307/2998573

Rabin, M., & Schrag, J. L. (1999). First Impressions Matter: A Model of Confirmatory Bias. *The Quarterly Journal of Economics*, *114*(1), 37–82. https://doi.org/10.1162/003355399555945

Schiffrin, Anya, and Stiglitz, J.E. (2020). Facebook Does Not Understand the Marketplace of Ideas. The Financial Times.

Sethi, R., & Yildiz, M. (2012). Public Disagreement. *American Economic Journal: Microeconomics*, *4*(3), 57–95. https://doi.org/10.1257/mic.4.3.57

Sethi, R., & Yildiz, M. (2016). Communication With Unknown Perspectives. *Econometrica*, *84*(6), 2029–2069. https://doi.org/10.3982/ECTA13320

Sobel, J. (2013). Giving and Receiving Advice. In D. Acemoglu, M. Arellano, & E. Dekel (Eds.), *Advances in Economics and Econometrics* (pp. 305–341). Cambridge University Press. https://doi.org/10.1017/CBO9781139060011.011

Stiglitz, J. E. (1977). Monopoly, Non-Linear Pricing and Imperfect Information: The Insurance Market. *The Review of Economic Studies*, *44*(3), 407–430. https://doi.org/10.2307/2296899

Stiglitz, J. E. (1984). Information, Screening and Welfare. *Bayesian Models in Economic Theory*, Marcel Boyer and Richard Khilstrom (eds.), Elsevier Science Publications, 209-239.

Stiglitz, J.E. (2012). *The Price of Inequality: How Today's Divided Society Endangers Our Future*. W.W. Norton.

Stiglitz, J.E. (2015). New Theoretical Perspectives on the Distribution of Income and Wealth among Individuals: Part IV: Land and Credit. NBER Working Paper No. 21192

Stiglitz, J.E. (2019*). People, Power, and Profits: A Progressive Capitalism for an Age of Discontent*. Penguin Books.

Svensson, Lars E. O. (2006) "Social Value of Public Information: Comment: Morris and Shin (2002) Is Actually Pro Transparency, Not Con." *American Economic Review*, 96(1), pp. 448-452.

Szeidl, Adam, and Ferenc Szucs. (2022). The Political Economy of Alternative Realities. Mimeo.

Taneva, I. (2019). Information Design. *American Economic Journal: Microeconomics*, *11*(4), 151–185. https://doi.org/10.1257/mic.20170351

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*(4), 297–323. https://doi.org/10.1007/BF00122574

Woodford, Michael. 2002. “Imperfect Common Knowledge and the Effects of Monetary Policy.” In Knowledge, Information, and Expectations in Modern Macroeconomics, edited by Philippe Aghion et al., 25–58. Princeton, NJ: Princeton University Press.