A WELFARE CRITERION FOR MODELS WITH DISTORTED BELIEFS*

Markus K. Brunnermeier
Alp Simsek
Wei Xiong

This article proposes a welfare criterion for economies in which agents have heterogeneously distorted beliefs. Instead of taking a stand on whose belief is correct, our criterion asserts that an allocation is belief-neutral efficient (inefficient) if it is efficient (inefficient) under any convex combination of agents' beliefs. Although this criterion gives an incomplete ranking of social allocations, it can identify positive- and negative-sum speculation driven by conflicting beliefs in a broad range of economic environments. *JEL* Codes: D61, D62, D84, G12, G14.

I. Introduction

The burgeoning behavioral finance and economics literature has identified a set of psychological biases that distort people's beliefs and decisions in various economic situations.[1] The presence of belief distortions stimulates normative analysis of welfare consequences of belief distortions. A standard approach taken by the literature is to assume that the social planner knows the objective belief measure and uses the objective measure to evaluate agents' welfare.[2] This approach, however, faces a major challenge

1. See Hirshleifer (2001), Barberis and Thaler (2003), and DellaVigna (2009) for extensive reviews of the literature.

2. For example, see Weyl (2007), Sandroni and Squintani (2007), Hassan and Mertens (2011), Gennaioli, Shleifer, and Vishny (2012), and Bianchi, Boz, and Mendoza (2012), and Spinnewijn (2014).

in implementation—which belief should the planner use? In many realistic situations, the planner does not observe the objective belief and faces the same difficulty as individuals do in discriminating different beliefs based on available data. Perhaps due to this challenge, many studies in the behavioral literature shy away from making any normative statement.

This article proposes a belief-neutral welfare criterion, which requires the planner to be sure of the presence of belief distortions by some agents but without having to precisely identify the objective belief. To illustrate the basic idea, we first consider a bet between two famous economic theorists, Joe and Bob.[3] One day, Joe and Bob argued over the contents of a pillow. Joe maintained that the pillow had a natural-fiber filling, and Bob thought a polyester filling was more likely. Joe assessed with probability 0.9 that the pillow had natural down and Bob assessed the probability at 0.1. They decided to construct a bet as follows: If the pillow had natural down, Bob would pay Joe $100, but if it had artificial down, Joe would pay Bob $100. They could discover the truth only by cutting the pillow open, which would destroy it. They agreed that the winner would replace the pillow at a cost of $50.[4] Clearly both Joe and Bob preferred the bet relative to no betting at all, as each expected to make a net profit of $35 after deducting the cost of replacing the pillow. This bet was desirable from each individual's perspective, and thus it Pareto dominated no betting under the standard Pareto principle. However, the outcome of the bet was worrisome—it led to a wealth transfer between Joe and Bob and a perfectly good pillow being destroyed.

Joe and Bob might have taken the bet for its entertainment value, which could have justified the cost of destroying the pillow. Another possible motive was that each bettor believed he would win and the other would lose. A planner could simply verify the reason by asking them. If the bet was motivated by a belief in winning, then at least one of them was overconfident, even though it was still difficult to tell who was overconfident. In this case, it is immediately obvious that the bet was a negative-sum game regardless of whose belief the planner uses to evaluate

---

3. See Kreps (2012, p. 193) for more details of the story.

4. We can also make the example more realistic by making the replacement cost of the pillow state-dependent, that is, the cost being $50 if it had natural down and $20 if it had artificial down. Our welfare analysis of the bet is robust to such a state-dependent replacement cost.

the social welfare. The resulting social loss is exactly the destroyed pillow.

In fact, the conflicting beliefs of Joe and Bob induced a form of externality. Despite knowing the bet would lead to the pillow being destroyed, each believed that he would win and the other party would lose. In this setting with conflicting beliefs, the meaning of "externality" needs to be broadened.[5] From Bob's perspective, his action causes an externality on Joe, even though Joe does not see it this way. From Joe's perspective, there is an analogous externality. The standard libertarian view does not restrict the choice of any individual if it does not cause a negative externality on others. We modify/extend this libertarian viewpoint to a setting with heterogeneous beliefs. In this setting, externality has to be evaluated under the belief of the individual whose choice causes the externality rather than under the belief of the individual who is exposed to the externality. In our pillow example, under the belief of either Joe or Bob, the negative externality on the other even exceeds his own gain. This negative sum serves as the basis for our welfare criterion.

To generalize the key insight of this example, we acknowledge the relevance of a set of reasonable beliefs and require efficiency to be robust across all of the reasonable beliefs. Our welfare criterion asserts a social choice to be belief-neutral (in)efficient if and only if it is (in)efficient under every reasonable belief. A key presumption of this criterion is that the planner is sure that some agents' beliefs are distorted.[6] Specifically, we accept all convex combinations of agents' beliefs as reasonable beliefs, as long as they are consistent with the commonly agreed upon aggregate statistics. We propose to use all of them to extend the two standard welfare analysis approaches—the expected social welfare approach and the Pareto efficiency approach.

The expected social welfare approach directly compares two social allocations x and y for a given welfare function. Our welfare criterion posits that y is belief-neutral inferior to x if the expected

---

5. One might disagree about this broadened use of the term *externality*—in fact, one of us does. Alternatively, one could coin a new term for a situation in which one benefits at the expense of others under one's own beliefs but not under the other's.

6. One can use sound nonchoice evidence to rule the presence of belief distortions in observed choices. See the recent contributions of Bernheim and Rangel (2007, 2009) and Koszegi and Rabin (2007) along this line.

total welfare from y is lower than that from x under every convex combination of the agents' beliefs. Let's go back to the bet between Joe and Bob. Suppose that the planner is sure that the bet was induced by belief distortions and that Joe and Bob were both risk-neutral. If the social planner assigns Joe and Bob equal weight in summing up their utilities in the social welfare function, it is clear that the bet is belief-neutral inferior to the status quo (no betting). This is because, regardless of which reasonable belief the social planner adopts to evaluate Joe's and Bob's expected utilities, the transfer of $100 between them has no impact on the expected social welfare, but destroying the pillow leads to a sure social loss of $50.

Without relying on a particular social welfare function, we can also adopt the Pareto dominance approach. Our criterion asserts that an allocation y is belief-neutral Pareto inefficient if, under every reasonable belief, there exists an alternative allocation y' that improves the expected utilities of some agents without hurting anyone else. Returning to the example, suppose that the planner adopts Joe's belief. Under this belief, the bet leads to an expected wealth transfer of $80 from Bob to Joe and the pillow's destruction. Alternatively, a direct transfer of $80 from Bob to Joe without the bet improves everyone's expected utility by saving them the cost of replacing the pillow. Similarly, under every convex combination of Joe's and Bob's beliefs, the planner can find a suitable (belief-measure dependent) transfer without the bet to strictly improve everyone's expected utility. Thus, the bet is belief-neutral inefficient with respect to any social welfare function that increases with agents' utilities.

In summary, without taking a stand on which belief was correct, the planner can categorically determine that the bet leads to an inefficient social outcome. The key is that the externality induced by the conflicting beliefs of Joe and Bob is uniformly negative under every reasonable belief. Of course, this feature may not always hold in a more general situation. For illustration, let us extend the bet. Suppose that Bob believed the pillow contained poisonous materials (instead of polyester) with 90% probability and that there is a social gain of $100 from removing a poisonous pillow from the public (instead of the $50 cost of replacing the pillow). The bet had a positive sum under Bob's belief but still had a negative sum under Joe's belief. Thus, it is neither belief-neutral efficient nor belief-neutral inefficient.

Despite its incompleteness, our belief-neutral criterion is able to identify negative and positive externalities from belief distortions in a variety of settings. Section 3 illustrates the potential applications using a series of examples, some of which are simplified versions of prominent models in the literature.

The first example extends and generalizes the pillow example in three ways. First, we make agents risk-averse. Trading induced by heterogeneous beliefs makes agents' consumption more volatile than their endowments and results in a negative-sum game in expected utility terms regardless of the planner's belief. Second, we add a hedging motive to the speculative motive, by making agents' endowments negatively correlated. In this case, a trade-off arises between the welfare gain from risk sharing and the welfare loss from speculative trading (see, e.g., Kubler and Schmedders 2012; Posner and Weyl 2013; Simsek 2013a). Third, we also allow for endogenous information acquisition along the lines of Grossman and Stiglitz (1980). In a fully rational model with homogeneous prior beliefs, agents' incentive to collect information is subdued because trading reveals part of the information to other traders for free. This positive externality leads to an underprovision of information that enhances socially beneficial hedging. Heterogeneous beliefs can provide a counterforce to this externality as the speculative motive provides additional incentive to acquire information, as informally discussed in Black (1986). Our criterion enables us to compare the hedging and information benefits with the costs of speculation due to belief heterogeneity.

Our next example investigates whether agents adequately self-insure, for example, whether motorists wear seat belts as a precaution for traffic accidents or banks retain capital as a precaution for financial losses. Suppose agents are optimistic about their own idiosyncratic risks although they know and agree on the aggregate risks, for example, each motorist knows the average accident probability but believes the accidents will happen to other motorists. In this setting, each agent chooses not to self-insure, even though she recognizes the benefit of self-insurance for the average agent. Although this example does not involve any negative externality, the restriction that agents agree on aggregate risks nevertheless makes it possible for our criterion to evaluate mandatory self-insurance policies observed in practice, such as seat belt laws or bank capital requirements.

The last two examples involve speculative bubbles. A number of studies (e.g., Harrison and Kreps 1978; Morris 1996; Scheinkman and Xiong 2003; Wu and Guo 2004; Hong, Scheinkman, and Xiong 2006; Hong and Sraer 2013) emphasize that the option to resell assets to future optimists can induce bubbles in asset prices. Our first example highlights how over-investment induced by price bubbles makes speculative trading a negative-sum game just like the bet between Joe and Bob (e.g., Gilchrist, Himmelberg, and Huberman 2005; Bolton, Scheinkman, and Xiong 2006; Panageas 2006). Our next example highlights that bubbles caused by heterogeneous beliefs can help overcome market breakdowns induced by the adverse-selection problem in lemons models (as in Akerlof 1970) and thus lead to a positive-sum game. Our criterion can also identify the consequent belief-neutral welfare gains.

Economists have long recognized that the standard Pareto criterion can lead to unappealing outcomes when agents hold conflicting beliefs. Early general equilibrium literature, for example, von Weizsäcker (1969), Dreze (1970), Starr (1973), Harris (1978), and Hammond (1981), noted that an allocation that is Pareto optimal in the usual sense might feature less than perfect risk sharing. In particular, it made a distinction between ex ante efficiency and various versions of ex post efficiency (with better risk-sharing properties). The independent decision theory literature, for example, Mongin (1997), Gilboa, Samet, and Schmeidler (2004), and Gilboa, Samuelson, and Schmeidler (2014), has also pointed out that the standard Pareto principle can be spurious when agents hold conflicting beliefs. Our contribution to these studies is to propose a belief-neutral criterion which circumvents the spurious unanimity problem under the premise that the planner is aware of the presence of belief distortions but unaware of the objective belief.

Another strand of the literature, for example, Stiglitz (1989), Summers and Summers (1989), and more recently Davila (2014), has emphasized the negative-sum nature of speculation in financial markets to make a case for a financial transaction tax. Our criterion and examples capture the costs as well as the benefits of speculation, therefore providing a framework for analyzing how speculative activities in financial markets should be regulated.

The article is organized as follows: Section II describes the welfare criterion in a generic setting. Section III provides a series of examples to demonstrate the capability of the criterion to

generate clear welfare ranking in a variety of environments with distorted beliefs. We conclude in Section IV. The technical proofs are provided in the Appendix.

II. THE WELFARE CRITERION

We introduce the welfare criterion in a generic setting with N agents, indexed by $i \in \{1, 2, \ldots, N\}$ and with $T + 1$ dates: $t = 0, 1, \ldots, T$. At time t, let s_t be the state of the economy, which summarizes the history of the economy up to t. Agent i's conditional transition probability at time t from state s_t to state s'_{t+1} at $t + 1$ is $\pi^i_t(s_t, s'_{t+1}) \geq 0$. The subscript t indicates that beliefs can be time-varying, and superscript i indicates that beliefs are potentially heterogeneous across agents. We summarize agent i's beliefs by $\Pi^i = \{\pi^i_t(s_t, s'_{t+1})\}_{t=0}^{T-1}$. We assume that agents consume only at the final time T. A social choice x represents a set of consumption allocations to all agents along the path of s_T: $x = \{x^i(s_T)\}$. Note that $x^i(s_T)$ is a vector of consumption to agent i. A feasible allocation satisfies the aggregate budget constraint at each final state.

Suppose that agent i has state-dependent utility function $u_i[s_T, x^i(s_T)]$ over the consumption stream $x^i(s_T)$. This function is strictly increasing and locally concave with respect to consumption. This utility specification is sufficiently general to capture the standard utility functions used in most economic models. Based on the utility specification and the agent's beliefs, his expected utility at time 0 is $E^i_0[u_i[s_T, x^i_T(s_T)]]$, where the superscript i denotes the expectation under agent i's beliefs. By building on expected utilities, our framework ignores preferences that feature ambiguity aversion.[7]

II.A. *Heterogeneous Beliefs*

We let agents hold different beliefs (i.e., $\Pi^i \neq \Pi^{i'}$ if $i \neq i'$) and assume the beliefs are common knowledge among the agents. Before we dive into welfare analysis, it is useful to sort out different sources of heterogeneous beliefs. Throughout our later analysis, we treat agents' beliefs as given. It is straightforward

---

7. Our later examples all use Markovian state structures and consumption at the final date T.

to think of the beliefs as outcomes of the agents' learning processes. Suppose that an unobservable variable π determines the probability of the tree moving up each period. Each agent has a prior belief about the distribution of π, observes some information about π in each period, and uses Bayes's rule to update his belief about π. Through this learning process, three sources may lead to heterogeneous beliefs among agents: (i) distortions in updating, (ii) different information, and (iii) different prior beliefs.

We emphasize distortions in updating as a key source of heterogeneous beliefs. A large branch of the academic literature highlights that people suffer from a range of well-established psychological biases, such as overconfidence, limited attention, representativeness bias, and conservatism in making financial decisions. See Hirshleifer (2001), Barberis and Thaler (2003), and DellaVigna (2009) for extensive reviews of the literature. These biases cause agents to react differently to information. In particular, overconfidence causes agents to exaggerate the precision of certain noisy signals and thus overreact to the signals. When agents overreact to different signals, they may end up with substantially different beliefs and, as a result, may speculate against each other.

The presence of belief distortions prompts welfare concerns. Some agents may be unaware of their belief distortions and, as a result, take actions that hurt their own and others' welfare. Thus, it is important that a social planner evaluates each agent's welfare by using the objective probability measure, which serves as the premise of our welfare criterion.

A second source of belief differences is asymmetric information. The well-known no-trade theorem (e.g., Aumann 1976; Milgrom and Stokey 1982; Sebenius and Geanakoplos 1983) shows that asymmetric information cannot cause rational agents with a common prior belief to hold common knowledge of heterogeneous beliefs or to trade with each other. This result motivates us to mostly ignore asymmetric information in our analysis, except in our example considered in Section III.D.

A third source of belief differences is heterogeneous prior beliefs. The decision theory literature that builds on Savage's (1954) notion of subjective probability treats beliefs separately for individual agents. As economics does not offer much guidance on how individuals form their prior beliefs, economists tend to agree that prior beliefs probably depend on an individual's background and experience. Morris (1995) summarizes a series of

arguments to advocate the view that rational agents may hold heterogeneous prior beliefs, just like heterogeneous risk preferences. In Brunnermeier and Parker (2005) and Brunnermeier, Gollier, and Parker (2007), heterogeneous prior beliefs arise endogenously from agents' anticipatory utility. In our analysis, we abstract away from agents' heterogeneous priors and instead focus on heterogeneous beliefs caused by distortions in updating.

II.B. *Welfare Analysis with Distorted Beliefs*

In the presence of distorted beliefs, it is important that the social planner uses an objective probability measure to evaluate agents' expected utilities in the welfare analysis. The challenge here is that the social planner may not observe the probability that drives economic uncertainty. Given the agents' different belief measures, whose measure is appropriate for welfare analysis? Is there an even more appropriate one outside of those used by the agents? We now introduce a belief-neutral welfare criterion.

Without taking a stand on which agent's belief is correct, we allow the planner to consider every belief from a set of reasonable beliefs. In our baseline analysis, this set contains all convex combinations of the agents' beliefs. Denote Π^h to be a convex combination of the agents' beliefs with weight $h = \{h^1, \ldots, h^N\}$:

$$\Pi^h = \sum_i h^i \Pi^i, \quad \text{where} \quad h^i \geq 0 \quad \text{and} \quad \sum h^i = 1.$$

In settings with multiple events, we further restrict the set of reasonable beliefs to satisfy commonly agreed-on aggregate statistics.[8] As we discuss in Section II.C, our criterion can be generalized to use alternative specifications for the set of reasonable beliefs.

The key contribution of our welfare criterion is that it allows for analysis of the efficiency of a social allocation according to all reasonable beliefs. Specifically, we propose to identify an allocation as inefficient (or efficient) if the social planner finds it inefficient (or efficient) under every reasonable probability measure Π^h that is commonly used to evaluate all agents' expected utilities. We can use two different approaches to implement our

---

8. The example in Section III.B clarifies this aggregate statistics restriction on reasonable beliefs.

proposal, one based on a given social welfare function and the other through the notion of Pareto efficiency. As is well known from standard economic theory, in the absence of belief distortions these two approaches are internally consistent. In particular, any Pareto-efficient social allocation corresponds to an optimal allocation that maximizes the agents' aggregate expected utilities under a set of nonnegative weights.

1. Belief-Neutral Social Welfare Criterion. The Bergsonian social welfare function is a sum of agents' expected utilities $\{E_0^h[u_i]\}$ (calculated according to a common measure Π^h) using a set of nonnegative weights $\{\lambda_i\}$:

$$W\big(E_0^h[u_1], E_0^h[u_2], \ldots, E_0^h[u_N]\big) = \sum_{i=1}^N \lambda_i E_0^h[u_i] = E_0^h\left[\sum_{i=1}^N \lambda_i u_i\right].$$

If the weights are all equal, it becomes the utilitarian social welfare function:[9]

$$W\big(E_0^h[u_1], E_0^h[u_2], \ldots, E_0^h[u_N]\big) = \sum_{i=1}^N E_0^h[u_i] = E_0^h\left[\sum_{i=1}^N u_i\right].$$

Based on a given welfare function, we can implement our criterion as follows.

DEFINITION 1. Consider two social allocations, x and y. If the expected social welfare of allocation x dominates that of allocation y for every reasonable probability measure Π^h,

$$W\big(E_0^h\big[u_1\big(s_T, x_T^1(s_T)\big)\big], \ldots, E_0^h\big[u_N\big(s_T, x_T^N(s_T)\big)\big]\big)$$
$$\geq W\big(E_0^h\big[u_1\big(s_T, y_T^1(s_T)\big)\big], \ldots, E_0^h\big[u_N\big(s_T, y_T^N(s_T)\big)\big]\big)$$

with the inequality holding strictly for at least one reasonable measure, then allocation x is belief-neutral superior to allocation y.

9. Given that these social welfare functions are linear and that the social planner uses the same probability measure to evaluate the expected utilities of all agents, the expected social welfare is independent of the order of aggregating welfare and computing expectations. In our analysis, we find it more convenient to first aggregate agents' welfare in each of the final states and then compare the expected social welfare under different probability measures.

To establish the superiority of one social allocation relative to another, a higher expected social welfare according to every convex combination of the agents' beliefs is required. This proposed belief-neutral superiority is a partial ordering of social allocations. Given two social allocations x and y, the allocation x might dominate y in one measure and y might dominate x in another measure. In such cases, we would say that x and y are incomparable.

Despite its incompleteness, this criterion is nevertheless useful in detecting negative-sum speculation driven by distorted beliefs. We apply this criterion to analyze the bet between Joe and Bob. Suppose that both Joe and Bob are risk neutral, $u_{Joe}(w) = w$ and $u_{Bob}(w) = w$, and that the social planner uses the utilitarian social welfare function for a reasonable belief:

$$W\big(E_0^h[u_{Joe}], E_0^h[u_{Bob}]\big) = E_0^h[w_{Joe} + w_{Bob}] = w_{Joe} + w_{Bob}.$$

It is obvious that without any betting, regardless of the probability measure the social planner adopts, the social welfare is simply the sum of Joe's and Bob's initial wealth. The bet causes a transfer of \$100 between them and the pillow's destruction. The money transfer has no effect on the social welfare regardless of its direction or the probability measure the social planner adopts to evaluate the welfare. However, replacing the pillow incurs a sure cost of \$50 and therefore makes the bet a negative-sum game for every reasonable, common probability measure used to evaluate Joe's and Bob's expected utilities. Thus, the status quo allocation is belief-neutral superior to the bet.

The utilitarian social welfare function assigns equal weights to all agents. If the social welfare function puts a sufficiently high weight on one agent, say, Joe, then we cannot directly compare the two allocations x and y. This is because, under Joe's belief, the bet increases his own expected utility and thus the social welfare relative to the status quo allocation. However, this may not be the case under Bob's belief. The second version of our criterion addresses this concern by generalizing the notion of Pareto efficiency and establishes that the bet is belief-neutral inefficient regardless of the choice of social welfare function.

2. *Belief-Neutral Pareto Efficiency.* The essence of Pareto efficiency is to determine whether there exists an alternative feasible allocation that improves the welfare (i.e., expected utility) of

some agents without hurting any other agent. If such an alternative exists, the allocation under evaluation is Pareto inefficient. We generalize this logic to environments with distorted beliefs to obtain a second implementation of our criterion.

DEFINITION 2. Consider a social allocation y. Suppose that, for every reasonable probability measure Π^h, there exists another (measure dependent) allocation y^h such that it improves some agents' expected utilities without reducing anyone's, that is, $\forall i$, $E_0^h\big[u_i\big(s_T, y_T^i(s_T)\big)\big] \leq E_0^h\big[u_i\big(s_T, y_T^{hi}(s_T)\big)\big]$ with the inequality holding strictly for at least one agent. In that case, allocation y is belief-neutral Pareto inefficient. In contrast, if, for every Π^h, there does not exist a dominating alternative, then allocation y is belief-neutral Pareto efficient.

If agents have common beliefs, that is, if Π^i is the same for each i, then the belief-neutral Pareto criterion coincides with the usual Pareto criterion. In the presence of distorted beliefs, as discussed before, the social planner uses a common probability measure from the set of reasonable measures to evaluate each agent's expected utility. The belief-neutral criterion then identifies an allocation as inefficient (or efficient) if it is Pareto inefficient (or efficient) under every reasonable measure.

Returning to the bet between Joe and Bob, we can show that the betting allocation, denoted y, is belief-neutral Pareto inefficient. Under Joe's belief, y is dominated by an alternative allocation, y^h, which keeps the pillow intact and simply transfers \$80 from Bob to Joe. This allocation improves both Joe's and Bob's expected utilities, under Joe's belief. Similarly, under Bob's belief, y is dominated by an alternative allocation that keeps the pillow intact and transfers \$80 from Joe to Bob. More generally, under every convex combination of Joe's and Bob's beliefs, there exists an appropriate direct transfer that improves the expected utilities of both Joe and Bob. The gain from such a transfer is due to saving the pillow from destruction. The bet is thus belief-neutral Pareto inefficient. It is also easy to see that the status quo allocation is belief-neutral efficient, as for every reasonable belief the planner cannot find a transfer to improve Joe's or Bob's welfare without hurting the other's. Taken together, the status quo allocation is on the belief-neutral Pareto-efficient frontier while the betting allocation is in the belief-neutral inefficient set.

Recall from the standard welfare theory (e.g., Mas-Colell, Whinston, and Green 1995, proposition 16.E.2) that each allocation on the Pareto frontier maximizes a linear social welfare function corresponding to some Pareto weights. This observation leads to the following result, which states belief-neutral Pareto inefficiency in terms of social welfare maximization.

PROPOSITION 1. Let X denote the set of all feasible allocations. Accordingly, an allocation, $x \in X$, is belief-neutral Pareto efficient (inefficient) if and only if, for every reasonable probability measure Π^h, there exists (does not exist) a set of Pareto weights $\{\lambda_i^h\}$ (with $\lambda_i^h > 0$ for all i and $\sum_i \lambda_i^h = 1$) such that

$$x \in \arg\max_{\hat{x} \in X} \sum_{i=1}^{N} \lambda_i^h E_0^h \Big[u_i \Big(s_T, \hat{x}_T^i(s_T) \Big) \Big].$$

Proposition 1 illustrates the relationship between the two versions of our criterion. Both versions consider all reasonable beliefs (i.e., convex combinations of agents' beliefs), which is the key characteristic of our approach. However, the welfare-function-based criterion fixes a particular social welfare function (e.g., a particular set of Pareto weights). By doing so, it enables us to compare allocations directly, for example, to say that the status quo allocation, x, is belief-neutral superior to the betting allocation, y. In contrast, the Pareto-efficiency version is more general because it considers not only all reasonable beliefs but also all social welfare functions (e.g., all possible Pareto weights). The cost of this generality is that the criterion does not provide direct comparisons between two allocations. Rather, it categorizes allocations into three sets: (i) those that are belief-neutral inefficient because they are inferior under every reasonable belief and every welfare function, (ii) those that are belief-neutral efficient because under every reasonable belief they are superior at least according to one welfare function, and (iii) those that are neither uniformly efficient nor uniformly inefficient across all reasonable beliefs.

II.C. Comments on the Criterion

1. Incompleteness. Our belief-neutral criterion requires the externality induced by agents' conflicting beliefs to be uniformly positive or negative across the set of reasonable beliefs. This

requirement is demanding and may lead to an incomplete ranking in some situations. To illustrate this incompleteness, we extend the bet between Joe and Bob. Suppose that Joe believed that the pillow was made of cotton with 90% probability as before, whereas Bob believed that the pillow contained poisonous materials with 90% probability. Again, they had to cut open the pillow to find out its content. If Joe was right, he would win $100 from Bob and pay $50 to replace the pillow. If Bob was right, he would win $100 from Joe. In addition, by removing the poisonous pillow from the public, there is a social benefit of $100, which we assume goes to Bob. For the sake of our discussion, we implement our criterion by using only a utilitarian social welfare function. It is easy to see that if the planner uses Joe's belief to evaluate the social welfare, the bet induces a negative sum. However, if the planner uses Bob's belief, the bet induces a positive sum due to the reward for discovering a poisonous pillow. Taken together, the bet is neither belief-neutral superior nor belief-neutral inferior relative to the status quo. This incomplete ranking reflects the belief-dependent cost and benefit of the bet in this case.

Furthermore, in this extended example, the sets of belief-neutral efficient and belief-neutral inefficient allocations are both empty. Given that these sets can be empty, our criterion instructs the planner to (i) choose a belief-neutral efficient allocation if it exists, (ii) avoid a belief-neutral inefficient allocation if there is any, and (iii) otherwise avoid any market intervention.

2. *Collective and Cautious Paternalism.* As our criterion ignores agents' preferences under their own beliefs, it is naturally paternalistic. That said, the criterion features a specific and disciplined form of paternalism—which might be called collective paternalism—as it is designed to detect inefficiencies (or efficiencies) based on disagreements between agents in a group setting.

More specifically, our criterion is not designed to analyze the inefficiencies induced by irrational behavior of an individual agent. Consider an agent who invests a large fraction of her wealth in a particular company's stock. This investment decision may appear inefficient to a conscientious observer who holds a more neutral view of the company's stock than the agent and who thus believes the agent should diversify her investment away from the company. However, the decision is optimal under the agent's beliefs. Without taking a stand on the beliefs of the agent

and the observer, our criterion cannot identify the agent's investment decision as efficient or inefficient in isolation.

On the other hand, in an equilibrium context, when one group of agents holds different beliefs than another group due to belief distortions, the trading between the two groups can make their consumption excessively volatile (as we discuss in Section III.A). Our criterion can identify the negative sum in expected utilities induced by trading without ruling a particular group's choice as inefficient.[10]

Even in group settings, our criterion is further disciplined by the requirement that the planner knows agents' beliefs are distorted. In practice, the planner might rule the presence of belief distortions using evidence from psychology or neuroscience (along the lines of Bernheim and Rangel 2007, 2009; Koszegi and Rabin 2007). Alternatively, the planner can account for other reasons for trade and obtain belief distortions as a reasonable residual. For example, consider the evidence suggesting that individuals invest considerably in their own companies or in professionally close stocks and that they tend to lose money on these investments (see Doskeland and Hvide 2011). In this context, risk-sharing can be ruled out because it would require investment in the opposite direction. Trade based on individual investors' informational advantage can also be ruled out because it would result in realized gains on average as opposed to losses. Hence, using this type of reasoning, a planner can conclude observed trades in this context are likely to be based on belief distortions.

3. Set of Reasonable Beliefs. In our baseline analysis, we take the set of reasonable beliefs as those that correspond to convex combinations of agents' beliefs, subject to the aggregate statistics commonly agreed by the agents. We view this as a reasonable benchmark for two reasons. First, this set is sufficiently large to include all of the extreme beliefs held by any agent in a given environment. As illustrated by the examples in the next section, this set includes the beliefs of the optimists who bid up asset

---

10. In this context, it is also useful to contrast our criterion with that offered by Bewley. In Bewley (2002), an individual decision maker holds several belief distributions and overcomes inertia only if the new choice dominates the status quo under all belief distributions. Bewley's theory shares our feature of belief neutrality, but analyzes a single agent's decision problem rather than evaluating the welfare of many decision makers, each with a different (but single) belief distribution.

prices and who take highly leveraged positions, as well as the beliefs of pessimists who are constrained by short-sales restrictions from directly participating in asset markets. Second, the set of convex combinations is also appropriate given our focus on detecting the inefficiencies due to belief disagreements as opposed to belief mistakes shared by all agents. If the objective belief is outside our reasonable set, then there might be some welfare losses—due to an irrational behavior common to all agents— that go undetected by our criterion. But our criterion will be useful even in these scenarios to detect part of the welfare losses that stem from negative-sum speculation.

That said, as Definitions 1 and 2 illustrate, our welfare criterion can also be used with more flexible specifications for the set of reasonable beliefs. We envision that depending on the application, the set of reasonable beliefs can be taken to be larger—or perhaps smaller—than our baseline specification. For instance, if the planner has a priori knowledge that the objective belief is likely to be in a particular set, then reasonable beliefs can be extended to include that set.[11] Conversely, if the planner has a priori knowledge that certain beliefs do not correspond to the objective belief, then those beliefs could be excluded from the reasonable set even if they are held by some agents.

4. Improving an Inefficient Allocation. In welfare analysis, the planner is often concerned not only with whether a social allocation is efficient but also with how to improve on an inefficient allocation. If the planner follows a specific welfare function, the first version of our belief-neutral criterion can be applied to address both of these issues because it can directly rank an allocation y against any alternative y'.

If the planner cannot rely on a specific welfare function, one might wonder whether the second version of our criterion, the belief-neutral Pareto criterion, can address the issue of how to improve on an inefficient allocation, say, y. This concern arises because, according to Definition 2, an alternative allocation y^h that dominates y may depend on the belief measure Π^h that the planner uses to evaluate y. In other words, the planner has to specify a particular belief Π^h to implement a belief-dependent

---

11. As shall become clear, in endowment-economy settings, our criterion becomes discerning even if the reasonable beliefs are extended to include any belief that assigns nonzero probability to all relevant states.

alternative allocation. However, as we will show in several of our examples, the belief-neutral Pareto-efficient frontier is often nonempty. When this is the case, we can indirectly rank y as inferior to any allocation on the belief-neutral Pareto-efficient frontier. This is also the case in our initial example of the bet between Bob and Joe, in which the status quo allocation is on the belief-neutral Pareto efficient frontier while the betting allocation is in the belief-neutral inefficient set. Thus, without relying on a particular welfare function, the belief-neutral Pareto criterion would nevertheless suggest that the status quo is preferred to the betting allocation.

5. *An Alternative Criterion*. It is useful to compare our welfare criterion to the no-betting Pareto-dominance criterion of Gilboa, Samuelson, and Schmeidler (2014; henceforth GSS). They propose to extend the Pareto criterion in the presence of heterogeneous subjective beliefs by defining a choice x to dominate another choice y based on two conditions. First, each agent's expected utility under her own beliefs from x is higher than or equal to that from y, which is the standard Pareto condition. Second, there exists one belief under which the expected utility of each agent from x is higher than or equal to that from y. This second condition is additional and requires the existence of a common belief to rationalize the efficiency of x.

In the example of Joe and Bob, there is no common belief that can rationalize the betting allocation. Hence, GSS's second condition prevents the bet from no-betting Pareto dominating the status quo. However, their first condition also prevents the status quo from no-betting Pareto dominating the bet—since the bet is desirable according to traders' own beliefs. As a result, the betting and the status quo allocations cannot be compared according to their no-betting Pareto dominance criterion. In contrast, our belief-neutral criterion identifies the betting allocation as being belief-neutral inferior.

The differences between the two criteria can be understood by considering the main premises behind our approach. First, we envision scenarios in which the planner knows that agents' beliefs are distorted. Consequently, unlike GSS, we ignore agents' expected utilities under their own, possibly distorted beliefs. Our second premise is that the planner does not know the objective belief. Consequently, we require the planner to vary the

common-belief measure across a large set of reasonable beliefs so that the resulting welfare ranking is robust. In contrast, GSS require the existence of a single common belief to rationalize the efficiency of an allocation. Despite the seemingly restrictive robustness requirement, our criterion is able to identify positive and negative externalities in the bet between Joe and Bob, as well as in many other examples discussed in the next section.

III. Examples

This section provides a series of examples to demonstrate that the simple welfare criterion we propose, despite its incompleteness, can produce a surprisingly sharp welfare ranking in a wide range of economic environments with heterogeneously distorted beliefs. The key is that the externality induced by conflicting beliefs in these models is often uniformly positive or negative across different beliefs. In the example of the bet between Joe and Bob, the negative externality is reflected in the destroyed pillow. More generally, similar negative externalities can emerge through excessive risk taking, overinvestments, and bankruptcy costs. Meanwhile, positive externalities can arise from alleviating the free-rider problem in information acquisition and overcoming market breakdowns induced by adverse selection. This section uses simple models to illustrate these different sources of externalities and demonstrates that our welfare criterion provides a clear welfare ranking in each case. The second example on self-insurance shows that our criterion can provide a welfare ranking even without any externality, provided that there are certain restrictions on agents' belief disagreements—as for example a commonly agreed-on aggregate statistic. Our analysis offers important policy implications. However, the purpose of our simple examples is to isolate particular externalities rather than cover all relevant features necessary for making specific policy recommendations.

III.A. Speculation, Hedging, and Informational Efficiency

Our first example focuses on excessive trading induced by distorted beliefs, an issue that often arises in policy discussions. We present this example in three stages to explore several closely related conceptual issues. In the first stage, agents have constant endowments, but trading induced by their belief disagreements

makes their consumption risky. As a result, trading is a negative-sum game in expected utility terms under any convex combination of the agents' beliefs. As in the pillow example, the welfare cost here stems from excessive risk taking. In the second stage, we introduce hedging motives by allowing agents' endowments to be risky and perfectly negatively correlated. Nevertheless, belief disagreements make the agents trade beyond simply offsetting their endowment risks. Our criterion can again identify the welfare loss induced by distorted beliefs. In the third stage, we introduce informational frictions regarding the state of the economy, which governs the agents' endowment risks. The presence of informational frictions makes costly information acquisition socially desirable. However, without heterogeneous beliefs, agents cannot fully appropriate the social value of their information acquisition. Distorted beliefs in this case can help overcome this free-rider problem and restore informational efficiency.

1. Speculative Motive. We start with a one-period endowment economy setting with two agents, A and B. Each agent invests at time $t = 0$ and consumes only at $t = 1$. Each agent has a constant endowment at $t = 1$ denoted by w. That is, there is neither aggregate nor idiosyncratic endowment risk. Each agent has an increasing and strictly concave utility function $u(c^i)$. The two agents hold heterogeneous beliefs about the state of the world, which takes two values, a and b. For now, since the endowment is constant, one may interpret the state as a sunspot. There is a single risky asset with payoff $V(a) = 1$ in state a and $V(b) = -1$ in state b. The asset is in zero net supply, like derivative contracts or bets. Agent A assigns a probability of $\pi^A \in (0, 1)$ to state a, while agent B assigns $\pi^B \in (0, 1)$. The difference in beliefs causes the agents to engage in speculative trades against each other.[12]

12. A large class of economic models analyzes trading between agents who hold heterogeneous beliefs regarding economic fundamentals and the impact of their trading on equilibrium asset price dynamics (e.g., Detemple and Murthy 1994; Kurz 1996; Zapatero 1998; Basak 2000; Buraschi and Jiltsov 2006; Jouini and Napp 2007; David 2008; Dumas, Kurshev, and Uppal 2009; Xiong and Yan 2010; Dumas, Lewis, and Osambela 2011). A key insight of these models is that trading induced by heterogeneous beliefs can lead to endogenous fluctuations in agents' wealth distribution, which in turn amplifies asset price volatility and induces time-varying risk premia. Although these models can capture important dynamics of asset prices and risk premia, researchers tend to avoid making any welfare statement due to the lack of a well-specified welfare criterion. This simple example

Suppose that the asset is traded at a price of p at $t = 0$ Agent i ($i \in \{A,B\}$) chooses k^i, the number of contracts, to maximize his strictly concave expected utility:

$$\max_{k^i} \pi^i u\big(w + k^i(1 - p)\big) + \big(1 - \pi^i\big)u\big(w - k^i(1 + p)\big).$$

In the Appendix, we formally derive each agent's optimal trading strategy and the resulting asset price p. It is intuitive that in the presence of belief disagreements, each agent takes a nonzero position in the asset. The trading turns the agents' constant consumption allocation in the status quo into a risky one. The resulting market equilibrium is inefficient by either version of our welfare criterion. Suppose the planner has a utilitarian welfare function, $W(u_A, u_B) = u_A + u_B$. For any probability measure that lies between agents' belief measures, it is socially optimal to maintain deterministic consumption.[13] If $\pi^A \neq \pi^B$, the status quo allocation is belief-neutral superior to the market equilibrium allocation. We also apply the second version of our criterion. Under any belief measure $\pi \in (0,1)$, the status quo with an appropriate (measure-dependent) transfer Pareto dominates the equilibrium allocation (again, because the latter is more volatile). Thus, the equilibrium allocation is belief-neutral inefficient.

2. Speculative and Hedging Motives. To make the example more realistic, we now introduce risky endowments. Agent A's endowment is $w - z$ in state a and $w + z$ in state b, while agent B's endowment is $w + z$ in state a and $w - z$ in state b. Without loss of generality, we assume $z \in (0, w)$. The negatively correlated endowments motivate the two agents to trade for hedging purposes, in addition to acting on the speculative motive discussed in the previous version.

Suppose that agent i ($i \in \{A, B\}$) takes on a position of k^i in the risky asset. If agents have common beliefs, $\pi^A = \pi^B = \pi$, they will trade the asset to fully hedge their endowment risk. As a result,

---

serves to illustrate that our criterion can potentially fill this gap, albeit without delivering any implication for asset price dynamics.

13. In fact, for this example, the set of reasonable beliefs can be extended to include any measure that assigns a probability $\pi \in (0, 1)$ to state a.

each agent consumes a constant amount w regardless of the state. However, when agents have different beliefs, $\pi^A \neq \pi^B$, their equilibrium consumption is risky. This is because their pursuit of speculative gains causes them to deviate from the optimal risk-sharing allocation. Like in the previous version, the market equilibrium is again belief-neutral Pareto inefficient.[14] Specifically, we formally prove in the Appendix that for any measure with $\pi \in (0,1)$, the equilibrium allocation is dominated by the optimal risk-sharing allocation with a certain transfer (which depends on the belief measure).

However, in this case, the status quo allocation with endowment risks is also belief-neutral inefficient. The welfare ranking between the status quo and the equilibrium allocation depends on the relative magnitude of the agents' endowment risk and their belief disagreement. For example, if the endowment risk is large and the belief disagreement is small, then the equilibrium allocation belief-neutral dominates the status quo. This sheds light on the ongoing debate over financial innovation, see, for example, Posner and Weyl (2013). Introducing new tradable securities on the one hand allows agents to hedge their risk and on the other hand opens room for welfare-reducing speculation.

3. Social Value of Information. Besides speculation and hedging, trading can also occur for informational reasons. Traders can collect information and make trades based on it. The trading reveals (part of) the information to all market participants. This information is socially desirable if it improves investment efficiency in physical projects or enhances risk sharing. However, each agent has an incentive to free ride on the costly information acquisition of others (e.g., Grossman and Stiglitz 1980). When agents are fully rational, the level of information acquisition may be suboptimally low. Distorted beliefs can help mitigate this inefficiency. We now extend the example to show that our criterion permits a welfare analysis of the social value of information in the presence of distorted beliefs.

---

14. Kubler and Schmedders (2012) and Simsek (2013a) analyze richer settings that feature a similar trade-off between hedging and risk sharing. Our welfare criterion is also useful for analyzing the inefficiency of speculative trading in these richer settings.

Consider a setting in which agents A and B live in an economy with two possible regimes that are characterized in the previous two stages. In regime 1, both agents have constant endowments w, as described in Section III.A.1. In regime 2, they have perfectly negatively correlated endowments $y = \{(y_a^A, y_b^A), (y_a^B, y_b^B)\} = \{(w - z, w + z), (w + z, w - z)\}$, as described in Section III.A.2. Let us assume that both agents agree that the two regimes are equally likely. At $t = 0$, one of the agents, say, agent A, can acquire information, which perfectly reveals the regime the economy is in, at a personal acquisition cost $c \geq 0$.[15] This acquisition cost is in utility terms and is deducted from the agent's utility from wealth.

As before, agents can trade a risky asset with payoff $V(a) = 1$ in state a and $V(b) = -1$ in state b. The agents have (possibly) heterogenous beliefs about probabilities of states a and b. For simplicity, we impose symmetry, that is, $\pi^A = 1 - \pi^B$. If $\pi^A = \frac{1}{2}$, beliefs are homogeneous. For explicit derivation, we assume that both agents have a logarithmic utility function over wealth, $u(W^i) = \ln(W^i)$.

Let us first consider the case with common beliefs $\pi^A = \pi^B$. In this case, agents trade only to hedge the risk embedded in their endowments. If they know the regime the economy is in, they can mutually diversify away their endowment risk in regime 2 by trading the risky asset. Otherwise, any trading position on the risky asset yields a risky consumption stream in either regime 1 or regime 2, or in both. Acquiring information then plays a critical role in resolving this situation. While the costly signal is observed only by the information acquirer, that is, agent A, trading perfectly reveals the private information to agent B as well. Specifically, if the signal reveals to agent A that the economy is in regime 2, agent A will initiate a trade with agent B; on the other hand, if the signal reveals to agent A that the economy is in regime 1, agent A is indifferent between trading or not trading. Thus, in the absence of any strategic behavior (which we assume), the trade initiated by agent A reveals his private signal to agent B.

15. Strictly speaking, there is a continuum of each type of agent. In this example, we abstract from the free-rider problem among type A agents in acquiring the information. One can think of a monopolistic information seller, who sells the information only to an investment club of type A agents at a fixed cost higher than any single agent can afford.

Despite being socially desirable, the costly information may be underprovided in a competitive equilibrium. This follows from the public good nature of information: both traders benefit from knowing the regime, but a single trader, agent A, bears the cost of acquiring the information. As in Grossman and Stiglitz (1980), information revealed through trading conveys a positive externality to the uninformed agent that cannot be fully captured by the information acquirer. This externality may lead to an (inefficient) underprovision of information—specifically, when the information acquisition cost c exceeds agent A's individual gain but not the social gain.

The particular depiction of the information acquisition problem changes once traders are endowed with distorted beliefs. Under distorted beliefs, agents trade not only for risk-sharing purposes but also for speculative reasons. Speculation may be welfare enhancing, as it could give agent A additional incentives to acquire the costly information. Specifically, knowing which regime the economy is in allows agent A to fully hedge his endowment risk in regime 2, which in turn frees up his risk-bearing capacity to take a greater speculative position against agent B based on their heterogeneous beliefs about states a and b.[16] This speculative motive for acquiring information thus mitigates the externality in acquiring information and may lead to a belief-neutral efficient private provision of information, provided that the distortions of beliefs are not sufficiently strong.

In the Appendix, we formally derive the subgame-perfect equilibrium of the model by first computing both agents' trading strategies at $t = 0$ while taking as given agent A's information acquisition decision, and then solving agent A's optimal information acquisition policy. It is intuitive that agent A chooses to acquire information if and only if the acquisition cost c is lower than a threshold $c^{\mathrm{eq}}(\pi^A)$, his private value of information. His value of information depends on the two agents' beliefs, π^A and $\pi^B = 1 - \pi^A$. By letting the planner use a utilitarian welfare function, $W(u_A, u_B) = u_A + u_B$, we also define two other threshold levels, $\overline{c}^{\mathrm{eff}}(\pi^A)$ and $\underline{c}^{\mathrm{inef}}(\pi^A)$, $\forall \pi^A \in [0, 1]$. If $c \leq \overline{c}^{\mathrm{eff}}(\pi^A)$, acquiring information is belief-neutral efficient, while if $c \geq \underline{c}^{\mathrm{inef}}(\pi^A)$, acquiring information is belief-neutral inefficient. These two

16. This feature is reminiscent of Simsek (2013a), in that by helping agents better hedge their endowment risks, financial innovations allow agents to speculate more based on their heterogeneous beliefs.

thresholds measure the social value of information. It turns out that due to the symmetry in the two agents' final wealth in this example, these two thresholds coincide: $\bar{c}^{\mathrm{eff}}(\pi^A) = \underline{c}^{\mathrm{inef}}(\pi^A)$, which we simply denote by c^{eff}.

Figure I plots the two thresholds for the information acquisition cost against agent A's belief π^A, based on the following parameter values: $w = 10$ and $z = 5$. As π^A deviates from 0.5, there is greater belief disagreement between agents A and B, since $\pi^B = 1 - \pi^A$. The solid line plots the private value of information $c^{\mathrm{eq}}(\pi^A)$, below which agent A chooses to acquire information in the market equilibrium. The dotted line depicts $c^{\mathrm{eff}}(\pi^A)$, below which acquiring information is belief-neutral efficient. The shapes of these lines are symmetric around $\pi^A = 0.5$ due to the symmetric structure in the two agents' beliefs.

At the benchmark level $\pi^A = 0.5$, agents A and B have the same (correct) beliefs about the probabilities of the two states. In this case, $c^{\mathrm{eff}}(0.5) > c^{\mathrm{eq}}(0.5)$, which reflects the underprovision of information in the equilibrium. This is because agent A alone pays for the information acquisition, whereas both agents A and B benefit from the information. In fact, $c^{\mathrm{eff}}(0.5)$ is exactly double $c^{\mathrm{eq}}(0.5)$ due to the symmetric structure of this example.

As π^A deviates from 0.5, $c^{\mathrm{eq}}(\pi^A)$ rises. This is because, as the two agents' belief disagreement increases, agent A perceives a greater profit from trading against agent B. This opportunity for increased speculation motivates agent A to acquire information at a larger cost. Thus, by raising c^{eq}, belief distortions mitigate the underprovision of information in the market equilibrium.

As π^A deviates from 0.5, $c^{\mathrm{eff}}(\pi^A)$ drops. That is, as the two agents' belief disagreement increases, the social value of information decreases. This is because the acquired information not only improves the sharing of endowment risks between the two agents, but also allows them to speculate more based on their disagreements. The latter effect makes both agents' final consumption more volatile and thus reduces the social value of information.

Based on the way $c^{\mathrm{eq}}(\pi^A)$ and $c^{\mathrm{eff}}(\pi^A)$ intersect each other, Figure I illustrates four different regions. In region I, the information acquisition cost c is lower than both $c^{\mathrm{eq}}(\pi^A)$ and $c^{\mathrm{eff}}(\pi^A)$ and, as a result, agent A acquires information in the equilibrium and the information acquisition is belief-neutral efficient. In region II, c is higher than $c^{\mathrm{eq}}(\pi^A)$ but lower than $c^{\mathrm{eff}}(\pi^A)$ and, as a result, there is no information acquisition and the lack of
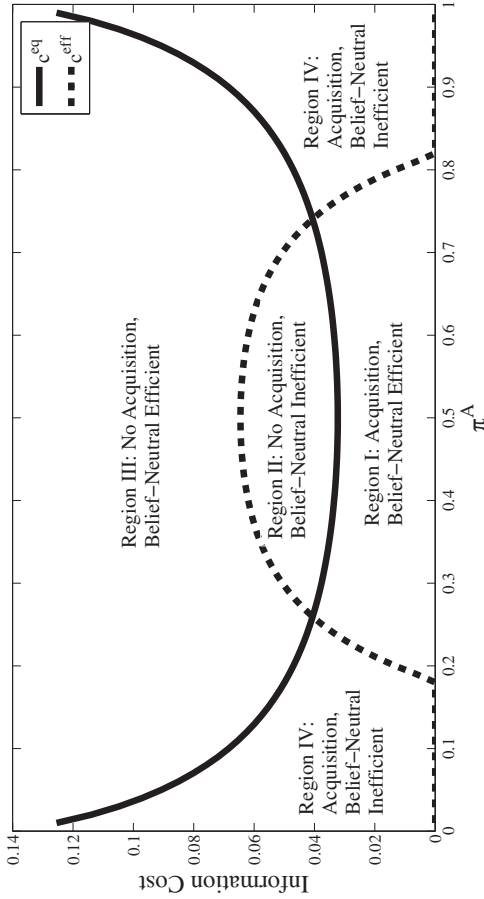
FIGURE I

Threshold Levels for Information Acquisition Cost

c^{eq} denotes the threshold for agent A to acquire information in equilibrium. c^{eff} refers to the threshold for acquiring information to be belief-neutral efficient. This plot uses the following parameter values: $w = 10$ and $z = 5$.

information acquisition is belief-neutral inefficient. As discussed earlier, a key insight of our analysis is that as the two agents' belief disagreement rises (i.e., π^A deviates further away from 0.5), this region narrows. In region III, c is higher than both $c^{\text{eq}}(\pi^A)$ and $c^{\text{eff}}(\pi^A)$. In this case, agent A does not acquire information and the lack of information acquisition is belief-neutral efficient. Finally, in region IV, c is higher than $c^{\text{eff}}(\pi^A)$ but lower than $c^{\text{eq}}(\pi^A)$. In this case, agent A acquires information and the information acquisition is belief-neutral inefficient.

The finance literature, for example, Black (1986), has long recognized the presence of noise trading induced by potential belief distortions of certain market participants as the key to resolving the free rider problem in information acquisition. However, there is little formal analysis of this issue due to the challenge in performing welfare analysis with the presence of distorted beliefs. This example shows that our criterion can help fill this gap in the literature.

III.B.   *Self-Insurance with Optimism*

It is well known that insurance markets might malfunction because of information asymmetries. However, these asymmetries cannot explain the failure of self-insurance arrangements that do not require market interaction. For instance, survey evidence suggests that less than 15% of motorists in the United States would wear seat belts voluntarily (see Williams and Lund 1986). This means that seat belt laws in the United States represent mandatory self-insurance. Regulations of this type are common also in other contexts. For example, financial regulation typically imposes on banks capital requirements that serve as insurance against potential losses.[17] We present a model of the failure of self-insurance arrangements based on optimism. Our criterion applied to this model can identify mandatory insurance allocations as belief-neutral superior to laissez-faire allocations. This example also highlights that our criterion can lead to clear welfare ranking even in the absence of any externality, as long as agents' belief disagreements satisfy certain restrictions.

---

17. These requirements are typically justified by moral hazard or fire-sale externalities. However, given that they are conceptually similar to seat belt laws, there might be room for an additional justification.

We develop the model in the context of seat belt laws. There are a large number of motorists denoted by $i \in I = \{1, \dots, N\}$. Each motorist i takes a precautionary action, $b_i \in \{0, 1\}$, where $b_i = 1$ corresponds to wearing a seat belt and $b_i = 0$ corresponds to not wearing one. After this decision, the motorist can be in one of two states denoted by $s_i \in \{0, 1\}$, where $s_i = 1$ corresponds to an accident and $s_i = 0$ corresponds to no accident. In case of an accident, the motorist suffers physical damage, which we capture with a monetary equivalent loss denoted by $d > 0$. For simplicity, suppose wearing a seat belt enables the motorist to completely avoid the damage. But wearing a seat belt is also inconvenient, which we capture with a monetary equivalent cost $c \in \{0, d\}$. We also assume the motorist is risk-neutral, so that the state utility function can be written as

$$u_i = -s_i(1-b_i)\,d - b_i c.$$

The aggregate state of the economy is given by $s = (s_i)_{i \in I} \in S$. Unlike our other examples, here the state involves multiple events. For simplicity, the economy features no aggregate uncertainty in the sense that exactly a fraction $\mu \in \{0, 1\}$ of motorists will have an accident, where μN is also an integer, so that

$$\sum_{i=1}^{N} s_i = \mu N \text{ for each } s \in S. \tag{1}$$

Importantly, all agents agree on this aggregate statistic, although each agent also believes her own state satisfies $s_i = 0$ with certainty. Put differently, motorists know and agree on the average accident probability (perhaps because they observe the historical accident statistics). Nonetheless, each motorist is optimistic in the sense that she believes these accidents will happen to other motorists.

Absent any policy requirement, each motorist chooses not to wear a seat belt so that the laissez-faire allocation features $b_i = 0$ for each $i \in I$. To evaluate welfare, consider the utilitarian social welfare function under any convex combination of motorists' beliefs. Since each motorist's belief satisfies the aggregate restriction in equation (1), so does any convex combination of their beliefs, which implies the utilitarian welfare,

$$E_0^h \left[\sum_{i=1}^{N} u_i \right] = -E_0^h \left[\sum_{i=1}^{N} s_i \right] d = -\mu N d.$$

Intuitively, the fraction μ of motorists will have an accident, which will lead to a social loss, $\mu N d$. Importantly, the inefficiency is detected by our belief-neutral criterion because motorists agree on the average accident probability.

Next consider a policy that makes it mandatory for all motorists to wear a seat belt so that $b_i = 1$ for each $i \in I$. The corresponding belief-neutral utilitarian welfare is

$$E_0^h \left[\sum_{i=1}^N u_i \right] = -Nc.$$

This policy features a different type of inefficiency. In particular, according to each motorist, at most $N - 1$ motorists—namely, motorists excluding himself—can have positive accident probability. Thus, each motorist believes that motorists in total are forced to wear more seat belts than is strictly necessary to reduce accident damage. The inefficiency is again detected by our belief-neutral criterion.

Comparing the two cases, the mandatory seat belt allocation is belief-neutral superior to the laissez-faire allocation if and only if

$$c < \mu d. \tag{2}$$

Intuitively, a seat belt is socially beneficial as long as its inconvenience cost, c, is smaller than its average damage reduction, μd. Thus, as long as the inconvenience cost is relatively small, our belief-neutral criterion provides a justification for the seat belt laws observed in practice.[18] Although there is not any externality in this example, the assumption that agents agree on the average accident probability—that restricts agents' belief disagreements—enables us to provide a clear welfare ranking.

18. The same point can be made using the Pareto version of our criterion, although the argument is more subtle than that derived from the utilitarian social welfare function. In this model, there are no belief-neutral Pareto-efficient allocations. This is because efficiency requires a subset of motorists to wear a seat belt, but which subset of motorists does so depends on the belief used for welfare analysis. That said, when condition (2) is satisfied, the laissez-faire allocation is belief-neutral Pareto dominated by the mandatory seat belt allocation combined with appropriate ex ante transfers. In this sense, the Pareto version of our criterion also favors the mandatory seat belt allocation over the laissez-faire allocation.

In the context of capital requirements for banks, the counterpart of an accident $s_i = 1$ can be suffering financial losses. The counterpart of wearing a seat belt $b_i = 1$ can be retaining sufficient equity capital (as opposed to paying out dividends) to absorb potential losses. The counterpart of the optimism assumption can be that banks know and agree on the average probability of suffering losses, but each bank believes these losses will be borne by other banks. Our analysis then suggests banks in a laissez-faire allocation will not retain sufficient capital to absorb losses. Moreover, as long as retaining capital is relatively cheap, mandatory bank capital requirements will generate a belief-neutral welfare improvement over the laissez-faire allocation. More broadly, our criterion offers a justification for mandatory self-insurance policies under the assumptions that agents agree on the aggregate risks, as in equation (1), and the cost of self-insurance is smaller than its average benefit, as in equation (2).

III.C. Bubble Models of Overinvestment

A segment of the literature emphasizes that when short sales are constrained, heterogeneous beliefs can lead to price bubbles as asset owners anticipate reselling their assets to other, more optimistic agents in the future (e.g., Harrison and Kreps 1978; Morris 1996; Scheinkman and Xiong 2003; Wu and Guo 2004; Hong, Scheinkman, and Xiong 2006; Hong and Sraer 2013). In these models, heterogeneous beliefs induce risk-neutral agents not only to trade against each other but also to overvalue assets. Overvaluation does not reduce social welfare by itself, because it is simply a welfare transfer across agents. However, overvaluation of equity can lead to firms' overinvestments (e.g., Bolton, Scheinkman, and Xiong 2006; Gilchrist, Himmelberg, and Huberman 2005; Panageas 2006), which reduce the total welfare of all investors. Our criterion can identify overinvestments independent of the belief used to evaluate firms' fundamental values.

We focus on a simple binomial setting with three dates (i.e., $t = 0, 1, 2$) and two risk-neutral agents (A and B). These agents trade the equity issued by a firm. The firm chooses its investment at time 0. Suppose that the firm's investment is cost-free, but the investment return has a decreasing return to scale. If the firm chooses to establish a production capacity of n units, the dollar return to per unit of capacity is determined by a binomial tree depicted in Figure II. There are three possible states
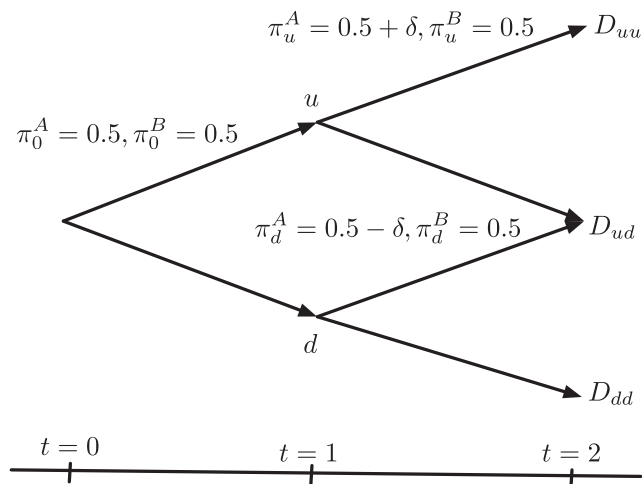
FIGURE II

Payoff and Belief Structure in the Bubble Model

(uu, ud, and dd) on $t = 2$. Suppose that the return per unit across the states at time 2 is

$$\tilde{R} = \{D_{uu}, D_{ud}, D_{dd}\} = \{R + 1 - n, R - n, R - 1 - n\},$$

where $R > 1$ is a constant. Due to the firm's decreasing return to scale, a larger investment scale n reduces the per unit return by n across all states at time 2. Suppose that the firm issues one share of equity for each unit of production capacity. The shares are equally distributed to A and B.

Before analyzing the firm's investment decision, we examine the market price of each share of equity. Figure II depicts the dynamics of the two agents' beliefs. We assume that the two agents have time-varying beliefs: They start with the same beliefs at time 0 but hold different beliefs at time 1:

$$\pi_0^A = \pi_0^B = 0.5, \pi_u^A = 0.5 + \delta > \pi_u^B = 0.5, \text{ and}$$
$$\pi_d^A = 0.5 - \delta < \pi_d^B = 0.5. \tag{3}$$

In particular, agent A becomes more optimistic than agent B in state u at time 1 and less optimistic in state d. The parameter $\delta > 0$ determines the two agents' belief dispersion in both states u

and d. It is straightforward to verify that, at $t = 0$, the two agents share the same expectation of the asset's final payoff:

$$E_0^A\left[\tilde{R}\right] = E_0^B\left[\tilde{R}\right] = R - n.$$

Following this literature, we assume that short-sale of the equity is not allowed. Accordingly, the fluctuations of the two agents' beliefs at $t = 1$ give an asset owner, who can be either A or B, an option to resell his holding to the other agent: more specifically, for agent A to sell to agent B in state d and for agent B to sell to agent A in state u. To obtain a bubble, we assume that each agent has sufficient cash to acquire the asset so that the competitive price is determined by the buyer's reservation value. It is straightforward to derive the following market price in state u: $p_u = R + \frac{1}{2} + \delta - n$, which is paid by agent A, and in state d: $p_d = R - \frac{1}{2} - n$, which is paid by agent B. By backward induction, both agents at time 0 value the asset by $p_0 = R + \frac{\delta}{2} - n$. Even though each agent's expectation of the asset payoff is $R - n$, their valuation of the asset is $R + \frac{\delta}{2} - n$. The difference is driven by the resale option, that is, the speculative motive to resell the asset to the other agent at a price higher than his own valuation at time 1. This resale option contributes a nonfundamental component to asset prices in the aforementioned bubble models.

We now analyze the firm's investment decision. Suppose the firm chooses its production capacity, n, to maximize its market value, given by $n \cdot p_0 = n \cdot \left(R - n + \frac{\delta}{2}\right)$. This is the appropriate objective function because the owners, agents A and B, agree on the firm's valuation at time 0. The firm's optimal investment level is then given by $n^* = \frac{1}{2}\left(R + \frac{\delta}{2}\right)$, which depends on δ, the magnitude of the two agents' belief dispersion at time 1.

Is this investment decision socially efficient? Suppose the planner uses the utilitarian social welfare function along with a convex combination of the two agents' beliefs, $\Pi^h = h\Pi^A + (1 - h)\Pi^B$, $\forall h \in (0, 1)$. Since both A and B are risk-neutral, the expected utilitarian social welfare is equal to the firm's expected final payoff, given by $n \cdot E^h\left[\tilde{R}\right] = n(R - n)$. This expression is maximized by choosing $n^{**} = \frac{1}{2}R < n^*$. This implies that the firm overinvests in the market equilibrium relative to the level that maximizes the expected utilitarian social welfare (or

the firm's long-run fundamental value) under any convex combination of the agents' beliefs.[19]

This result does not need to rely on any social welfare function because the market equilibrium is, in fact, belief-neutral Pareto inefficient. In particular, it can be checked that for any reasonable belief, Π^h, the market equilibrium with investment level n^* is Pareto dominated by an alternative allocation with investment $n^{**} < n^*$ combined with some initial transfer, $T \in [-n^* (R - n^*), n^* (R - n^*)]$, from agent B to agent A.

The driving force behind the inefficient overinvestment is exactly the value of the resale option in the firm's time 0 market valuation. Anticipating the possibility of reselling the share to the other agent at time 1 at a profit, each agent over-values the share at time 0 relative to his own expectation of the share's long-run fundamental value. This, in turn, induces the firm to overinvest. Note that each agent recognizes that this level of investment reduces the firm's long-run value. However, each agent also thinks that these losses will be borne by the other agent. A negative externality emerges just like in the bet between Joe and Bob. Consistent with this overinvestment example, Gilchrist, Himmelberg, and Huberman (2005) provide evidence that firms tend to increase investment in response to increased heterogeneous beliefs proxied by dispersion in analysts' earnings forecasts.

III.D. Benefits of Speculation in Lemons Models

The previous example shows that overinvestment in heterogeneous beliefs-induced bubble models leads to belief-neutral welfare losses. However, speculation and bubbles induced by heterogeneously distorted beliefs can also be beneficial. Among other things, bubbles help overcome market breakdown in "lemons" models caused by adverse selection (as in Akerlof 1970). This section illustrates this point by introducing heterogeneous beliefs into a recent model of Tirole (2012). Also see Morris (1994) for a model in which heterogeneous beliefs help break the no-trade

19. Given the presence of the firm's investment decision, it is important to restrict the set of reasonable beliefs to the convex combinations of agents' beliefs. This is because a measure outside the convex combinations of agents' beliefs would imply that the agents' aggregate belief is biased and thus rule the firm's investment decision in the equilibrium as inefficient, even in the absence of any belief dispersion between the two agents. As stated previously, analyzing inefficiencies associated with agents' aggregate biases is not our focus.

theorem and Zhuk (2012) for a model in which bubbles induced by heterogeneous beliefs help overcome the information externalities among firms.

The model of Tirole (2012) considers a firm that attempts to finance a new investment project by selling its legacy asset. However, the firm is asymmetrically informed about the payoff from the legacy asset, which creates a lemons problem. As in Akerlof (1970), the equilibrium features a low price and reduced trade and, in some extreme cases, a complete market breakdown. We show that bubbles induced by heterogeneous beliefs mitigate the lemons problem by allowing the firm to sell its asset and invest in the new project even if the quality of its legacy asset is relatively high. Our criterion can detect the consequent welfare gain.

Consider a seller who has access to a new project that costs I and generates a payoff of $I + G$. The payoff of the project is not pledgable (that is, it accrues to the seller but cannot be promised to others). Thus, the seller needs to finance the project by selling a legacy asset that is pledgable. This asset returns $\overline{R}$ with probability θ, and 0 otherwise. The probability, θ, itself is uniformly distributed over [0,1]. The prior value of the pledgable asset exceeds the investment cost, $p^{prior} \equiv \frac{\overline{R}}{2} > I$, so that the project is always financed in a constrained efficient allocation.

The key friction is that the seller is asymmetrically informed about the success probability of the legacy asset. In particular, the seller receives a signal and fully learns θ, while potential buyers continue to believe that θ is distributed according to the uniform prior. The rest of the section analyzes the effect of this friction on the efficiency of the equilibrium allocation with and without heterogeneously distorted beliefs. Suppose also that $G < \frac{\overline{R}}{2}$, which rules out the extreme case in which the seller is always able to finance the project despite having asymmetric information.

First, consider the benchmark without distorted beliefs among potential buyers. Let p^* denote the equilibrium asset price. If $p^* < I$, then there is no trade because the seller is unable to finance the new project by selling the legacy asset. If $p^* > I$, then a trade is possible. In particular, the seller will sell the asset only if $\theta\overline{R} \leq p^* + G$. In a competitive equilibrium, the buyers break even, which implies $p^* = \overline{R}E\left[\theta | \theta \leq \frac{p^* + G}{\overline{R}}\right]$. Solving this further gives the equilibrium price $p^* = G$. It follows that

there is no trade when $G < G^* = I$. When $G \geq G^*$, the seller will sell the asset only if $\theta < \theta^*$, and she will do so at a price p^*, where

$$\theta^* = \frac{2G}{\overline{R}} < 1 \text{ and } p^* = G < p^{prior}.$$

In particular, the adverse selection induced by the asymmetric information between buyers and the seller reduces the level of asset trading and the asset price. Intuitively, sellers with low-quality assets (lemons) exert a negative externality on sellers with higher-quality assets. In some cases (i.e., $G < I$), there is a complete market breakdown.

To formally discuss social welfare, we consider (as in Tirole 2012) the ex ante utilitarian social welfare function, that is, the sum of the seller's and buyers' expected utilities under the prior distribution for θ. Since the trading profits represent a pure transfer between the seller and buyers, the ex ante social welfare is simply

$$E\Big[\tilde{R}\theta + I_{\{G>G^*,\theta<\theta^*\}}G\Big] < \frac{\overline{R}}{2} + G.$$

Here, $I_{\{G>G^*,\theta<\theta^*\}}$ is an indicator function for whether the seller manages to invest in the project, and the inequality follows since there is investment with probability strictly less than 1. In contrast, an alternative (feasible) allocation that always transfers the asset from the seller to buyers at price p^{prior} ensures that the project is always financed and the social welfare is $\frac{\overline{R}}{2} + G$. Hence, the competitive equilibrium is (constrained) Pareto inefficient.

Next, we consider the case of buyers holding heterogeneously distorted beliefs regarding the asset return. Suppose that the asset return in the event of success is random and independent of the asset's success. We denote it by $\tilde{R}$ and assume that it can take two possible values, $\overline{R} + 1$ and $\overline{R} - 1$. The seller believes the probability of $\tilde{R} = \overline{R} + 1$ is 0.5. There are two groups of risk-neutral buyers for the asset. One group believes the probability of $\tilde{R} = \overline{R} + 1$ is 1, and the other group believes the probability is 0. Suppose that no one can short-sell the asset and each group has sufficient cash to acquire the asset. As in the previous example, buyers in the optimistic group acquire the asset and bid up its price to their expectation of the asset payoff. A key feature of the model is that the asset overvaluation induced by agents'

heterogeneous beliefs (as in Miller 1977) helps overcome the lemons problem. To see this most starkly, suppose $G > \frac{\overline{R}-1}{2}$ (while continuing to assume $G < \frac{\overline{R}}{2}$). Under this assumption, it can be seen that the seller chooses to sell and finance the project regardless of θ. The optimistic buyers break even only if $p = (\overline{R} + 1) E[\theta] = \frac{\overline{R}+1}{2}$. At this price, the seller in turn finds it optimal to sell because $\theta \overline{R} \leq \overline{R} < p + G$, where the last inequality follows since $G > \frac{\overline{R}-1}{2}$. Consequently, unlike the earlier case (for the same parameters), the competitive equilibrium with belief heterogeneity features trade and investment with probability 1.

We can apply our welfare criterion to show that the equilibrium with belief heterogeneity is, in fact, belief-neutral efficient. To see this, let Π^h denote a probability measure, which assigns probability $h \in [0,1]$ to $\tilde{R} = \overline{R} + 1$ and is a convex combination of all buyers' beliefs. The ex ante social welfare under this belief can be written as

$$E^h\Big[\tilde{R}\theta + G\Big] = E^h[\overline{R}]\frac{1}{2} + G,$$

since the project is invested with probability 1. As this expression illustrates, regardless of the probability measure, the ex ante welfare is at its highest possible level. This is because there is no disagreement about G, the gains from undertaking the project. This in turn implies that the equilibrium is belief-neutral efficient. Thus, speculation induced by heterogeneous beliefs mitigates the lemons problem and leads to belief-neutral welfare gains.

III.E. Other Examples

Our criterion is also applicable in other settings with distorted beliefs. To save space, we mention them only briefly and refer interested readers to our NBER working paper for more details.

A growing literature builds on agents' heterogeneous beliefs to analyze leverage cycles (e.g., Geanakoplos 2003, 2010; Fostel and Geanakoplos 2008; Cao 2010; Shen, Yan, and Zhang 2011; He and Xiong 2012; Simsek 2013b). The key feature of those models is that optimism can motivate cash-constrained optimists to use collateralized short-term debt to finance their asset acquisition. The leverage initially fuels the price boom but

later forces the optimists to deleverage after bad shocks, resulting in a leverage cycle. This framework nicely integrates the optimists' leverage cycle with the asset-price cycle. Both cycles are important for understanding various historical episodes of financial crises, including the recent one. To use this framework to analyze relevant policy issues (such as regulation over financial institutions' leverage), it is important to discuss welfare implications. Our criterion can generate a clear welfare ranking in this framework. The key insight is that overoptimism causes optimists to use excessive leverage in asset acquisition despite the possibility of incurring bankruptcy costs in the future. Bankruptcy costs make the excessive leverage a negative-sum game between optimistic buyers and pessimistic creditors.

In macroeconomic models, belief disagreements can also distort aggregate investment through individuals' consumption and savings decision; see, for example Sims (2008). Belief disagreements cause individuals to perceive greater expected returns from their investments. This affects their savings decisions in the same way an increase in the real interest rate does. It creates not only a substitution effect, which tends to increase savings, but also an income effect, which tends to increase current consumption and thus reduce savings. Depending on which effect dominates, individuals might save too much or too little relative to a homogeneous beliefs benchmark. The net saving in turn leads to over- or underinvestment. Our criterion can also help detect these types of inefficiencies.

IV. CONCLUSION

This article proposes a belief-neutral welfare criterion for models in which agents have heterogeneously distorted beliefs. The criterion builds on the premise that a planner is aware of belief distortions by some agents but cannot differentiate whose beliefs are distorted. The criterion rules that an allocation is belief-neutral efficient (inefficient) if it is efficient (inefficient) under any convex combination of the agents' beliefs. We can implement this criterion either through a given social welfare function or the notion of Pareto efficiency. Although this criterion gives an incomplete welfare ranking, it is nevertheless useful in identifying negative-sum or

positive-sum speculation. Through a series of examples, we show that this criterion is capable of identifying welfare gains/losses in a wide range of economic environments with heterogeneously distorted beliefs.

I. APPENDIX: TECHNICAL DERIVATIONS

A.I. *Technical Derivation for Section III.A.1*

The first-order condition to agent i's utility maximization implies

$$(1-p)\pi^i u'(w + k^i(1-p)) = p(1-\pi^i) u'(w - k^i(1+p)).$$

The market-clearing condition requires that $k^A + k^B = 0$. The standard results hold that there is a market equilibrium allocation, $\{k^A, k^B, p\}$, which solves each agent's optimality condition and the market-clearing condition.

First, we establish that, if $\pi^A \neq \pi^B$, the two agents will take a nonzero position in the contract. The first-order condition implies that

$$\frac{\pi^A}{1-\pi^A}\frac{u'(w + k^A(1-p))}{u'(w - k^A p)} = \frac{\pi^B}{1-\pi^B}\frac{u'(w + k^B(1-p))}{u'(w - k^B p)}.$$

Suppose that $k^A = k^B = 0$. Then, we must have $\frac{\pi^A}{1-\pi^A} = \frac{\pi^B}{1-\pi^B}$, which contradicts $\pi^A \neq \pi^B$. Thus, k^A and k^B cannot both be zero, which in turn implies that both are nonzero.

We now prove that if the social planner has the utilitarian welfare function $W(u_A, u_B) = u_A + u_B$, then the status quo allocation $y = \left\{(y_a^i, y_b^i) \equiv (w, w)\right\}_{i \in \{A,B\}}$ is belief-neutral superior to the market equilibrium allocation:

$$x = \left\{(x_a^i, x_b^i)\right\}_{i \in \{A,B\}} = \left\{(w + k^i(1-p), w - k^i(1+p))\right\}_{i \in \{A,B\}}.$$

Consider any measure with $\pi \in [\pi^B, \pi^A]$. The agents' utilitarian social welfare in the market equilibrium is given by

$$U^h = \pi\left[u(w + k^A(1-p)) + u(w - k^A(1-p))\right]$$
$$+ (1-\pi)\left[u(w - k^A p) + u(w + k^A p)\right].$$

The strict concavity of $u(\cdot)$ implies that

$$u\big(w + k^A(1 - p)\big) + u\big(w - k^A(1 - p)\big) < 2u(w),$$

$$u\big(w - k^A p\big) + u\big(w + k^A p\big) < 2u(w).$$

Thus, $U^h < \pi \cdot 2u(w) + (1 - \pi) \cdot 2u(w) = 2u(w)$, which is the utilitarian social welfare under the status quo. This proves that the status quo allocation is belief-neutral superior to the market equilibrium allocation.

Next, we show that for any measure with $\pi \in (0, 1)$, the equilibrium allocation is Pareto dominated by the status quo allocation with a certain transfer $T^\pi \in [-w, w]$, which leads to the following allocation: $y(T^\pi) = \big\{(w + T^\pi, w + T^\pi), (w - T^\pi, w - T^\pi)\big\}$. Consider each agent's certainty-equivalent wealth, $w^{i,eq}$, given by the solution to

$$u\big(w^{i,eq}\big) = \pi u\left(w + k^i(1 - p)\right) + (1 - \pi) u\left(w - k^i p\right), \forall i \in \{A, B\}.$$

The strict concavity of $u(\cdot)$ (along with the fact that $k^i \neq 0$) implies that

$$u\big(w^{i,eq}\big) < u\big(\pi(w + k^i(1 - p)) + (1 - \pi)(w - k^i p)\big).$$

Since $u(\cdot)$ is strictly increasing, this further implies

$$w^{i,eq} < \pi\big(w + k^i(1 - p)\big) + (1 - \pi)(w - k^i p), \forall i \in \{A, B\}.$$

Adding these inequalities and using market clearing, $k^A + k^B = 0$, we have $w^{A,eq} + w^{B,eq} < 1$. It follows that the status quo with an appropriate transfer Pareto dominates the equilibrium.

A.II. *Technical Derivation for Section III.A.2*

Suppose that agent i ($i \in \{A, B\}$) takes on a position of k^i in the risky asset. The position is characterized by the following first-order condition:

$$(1 - p)\,\pi^A\,u'\big(w - z + k^A(1 - p)\big) = p\big(1 - \pi^A\big) u'\big(w + z - k^A(1 + p)\big).$$

This is also a similar condition for agent B. In equilibrium, the market-clearing condition is $k^A + k^B = 0$.

Recall that the optimal risk-sharing trade is given by $k^{*A} = z$ and $k^{*B} = -z$. An analysis similar to the previous proof shows that, when $\pi^A \neq \pi^B$, agents deviate from the optimal risk-sharing trade, that is, $k^A \neq k^{*A}$. Next, fix a belief, π, and consider each agent's certainty-equivalent wealth under this belief given by the

solution to

$$u\big(w^{A,eq}(e,\pi)\big) = \pi u\big(w - z + k^A(1-p)\big)$$
$$+ (1-\pi)u\big(w + z - k^A(1+p)\big),$$

$$u\big(w^{B,eq}(e,\pi)\big) = \pi u\big(w + z + k^B(1-p)\big)$$
$$+ (1-\pi)u\big(w - z - k^B(1+p)\big).$$

Since $k^A \neq k^{*A}$, an agent's consumption is not constant across the states. Using the strict concavity of $u(\cdot)$, we obtain

$$w^{A,eq}(e,\pi) + w^{B,eq}(e,\pi) < 2w.$$

On the other hand, the status quo allocation, combined with the optimal risk-sharing trade, (k^{*A}, k^{*B}), gives each agent a constant consumption of w. It follows that, under any belief π, this allocation, combined with an appropriate transfer T^π, Pareto dominates the equilibrium allocation.

A.III. *Technical Derivation for Section III.A.3*

To solve the subgame-perfect equilibrium of the model, we first compute both agents' trading strategies at $t = 0$, while taking as given agent A's information acquisition decision, and then solve agent A's optimal information acquisition policy.

Consider the case in which agent A acquires the information at $t = 0$. Recall that agent A's trading position perfectly reveals the informational content of his private signal to agent B. Thus, both agents have the same information set at the time of choosing their optimal trading strategies. Let $W^i(S, s)$ denote the endowment of agent i in regime S and state s, where $S \in \{1, 2\}$ and $s \in \{a, b\}$. In view of log utility, the agent's problem can be written as

$$U_i^e(S) = \max_{k^i(S) \in \mathbb{R}} \big\{ \pi^i \ln\big[W^i(S, a) + k^i(S)\big] + (1 - \pi^i) \ln\big[W^i(S, b) - k^i(S)\big] \big\}.$$

$$(4)$$

The first-order condition for $k^i(S)$ is given by

$$\pi^i \frac{1}{W^i(S, a) + k^i(S)} - \big(1 - \pi^i\big) \frac{1}{W^i(S, b) - k^i(S)} = 0.$$

Note that, due to symmetry, the market-clearing condition $k^A(S) + k^B(S) = 0$ is automatically satisfied.

Consider now the case in which agent A does not acquire information at $t = 0$. As before, the agents' problems are symmetric. Specifically, agent i's problem is given by

$$U_i^e(N) = \max_{k^i(N) \in \mathbb{R}} \frac{1}{2} \Big[\pi^i \ln \big[W^i(1,a) + k^i \big] + (1 - \pi^i) \ln \big[W^i(1,b) - k^i \big]$$

$$+ \pi^i \ln \big[W^i(2,a) + k^i \big] + (1 - \pi^i) \ln \big[W^i(2,b) - k^i \big] \Big].$$

(5)

The first-order condition for $k^i(N)$ is given by

$$\pi^i \frac{1}{W^i(1,a) + k^i} - (1 - \pi^i) \frac{1}{W^i(1,b) - k^i}$$

$$+ \pi^i \frac{1}{W^i(2,a) + k^i} - (1 - \pi^i) \frac{1}{W^i(1,a) - k^i} = 0.$$

As in the previous case, note that the market-clearing condition $k^A(N) + k^B(N) = 0$ is automatically satisfied.

Agent A acquires information if and only if

$$\frac{1}{2} U_A^e(1) + \frac{1}{2} U_A^e(2) - c \geq U_A^e(N),$$

where $U_A^e(N)$, given in equation (5), is agent A's expected utility by not acquiring the information and $U_A^e(S)$, given in equation (4), is his expected utility conditional on acquiring the signal and the signal reveals regime S. This condition is equivalent to

$$c \leq c^{eq}(\pi^A) \equiv \frac{1}{2} U_A^e(1) + \frac{1}{2} U_A^e(2) - U_A^e(N).$$

(6)

Here, $c^{eq}(\pi^A)$ denotes the cost threshold below which agent A chooses to acquire information, characterized by equations (4) and (5).

Having solved for agent A's information acquisition policy, we now apply our welfare criterion to determine whether the private provision or lack of provision of information is either (i) belief-neutral efficient, (ii) belief-neutral inefficient, or (iii) neither belief-neutral efficient nor inefficient.

The set of reasonable beliefs is given by

$$B_R(\pi^A) = \big[\min\{\pi^A, 1 - \pi^A\}, \max\{\pi^A, 1 - \pi^A\}\big].$$

For any reasonable belief $\Pr(a) \in B_R$, the social welfare obtained when acquiring information, $W(I)$, is given by

$$W(I) = \sum_{S=1}^{2} \sum_{s=a}^{b} \sum_{i=A}^{B} \frac{1}{2}\Pr(s)\big\{\ln\big[W^i(S, s) + V(s)k^i(S)\big] - cI_{(i=A)}\big\},$$

while that obtained when no information is acquired, $W(N)$, is given by

$$W(N) = \sum_{S=1}^{2} \sum_{s=a}^{b} \sum_{i=A}^{B} \frac{1}{2}\Pr(s)\ln\big[W^i(S, s) + V(s)k^i\big].$$

Define the belief-neutral-efficient cost threshold $\overline{c}^{\mathrm{eff}}(\pi^A)$ as

$$\overline{c}^{\mathrm{eff}}(\pi^A) \equiv \sup\big\{c \geq 0 : W(I) \geq W(N), \forall \Pr(a) \in B_R(\pi^A)\big\}$$

and the belief-neutral-inefficient cost threshold $\underline{c}^{\mathrm{inef}}(\pi^A)$ as

$$\underline{c}^{\mathrm{inef}}(\pi^A) \equiv \inf\big\{c \geq 0 : W(I) \leq W(N), \forall \Pr(a) \in B_R(\pi^A)\big\}.$$

We now characterize $\overline{c}^{\mathrm{eff}}(\pi^A)$ and $\underline{c}^{\mathrm{inef}}(\pi^A)$, and prove that $\overline{c}^{\mathrm{eff}}(\pi^A) = \underline{c}^{\mathrm{inef}}(\pi^A), \forall \pi^A \in [0, 1]$. Define $x^i(S, s)$ and $y^i(S, s)$ as trader i's final wealth in state (S, s) when trader A acquires and does not acquire information, respectively:

$$x^i(S, s) \equiv W^i(S, s) + V(s)\,k^i(S)$$

and

$$y^i(S, s) \equiv W^i(S, s) + V(s)\,k^i(N).$$

From the market-clearing condition, it follows that

	$S = 1$		$S = 2$	
	$s = a$	$s = b$	$s = a$	$s = b$
$x^A(S, s)$	$w + k^A(1)$	$w - k^A(1)$	$w - z + k^A(2)$	$w + z - k^A(2)$
$x^B(S, s)$	$w - k^A(1)$	$w + k^A(1)$	$w + z - k^A(2)$	$w - z + k^A(2)$
$y^A(S, s)$	$w + k^A(N)$	$w - k^A(N)$	$w - z + k^A(N)$	$w + z - k^A(N)$
$y^B(S, s)$	$w - k^A(N)$	$w + k^A(N)$	$w + z - k^A(N)$	$w - z + k^A(N)$

It follows directly from the symmetry in the agents' payoffs that both $W(I)$ and $W(N)$ are independent of the belief that the planner uses to evaluate the social welfare:

$$W(I) = \sum_{S=1}^{2} \frac{1}{2}\{\ln[x^A(S,a)] + \ln[x^B(S,a)]\} - c$$

and

$$W(N) = \sum_{S=1}^{2} \frac{1}{2}\{\ln[y^A(S,a)] + \ln[y^B(S,a)]\}.$$

Accordingly, we have

$$\bar{c}^{\text{eff}}(\pi^A) = \underline{c}^{\text{inef}}(\pi^A)$$

$$= \sum_{S=1}^{2} \frac{1}{2}\{\ln[x^A(S,a)] + \ln[x^B(S,a)] - \ln[y^A(S,a)] - \ln[y^B(S,a)]\},$$

$$(7)$$

completing the characterization. The main text compares the cost thresholds characterized in equations (6) and (7) to assess the belief-neutral efficiency of agent A's private information acquisition decision.

PRINCETON UNIVERSITY AND NBER
MASSACHUSETTS INSTITUTE OF TECHNOLOGY AND NBER
PRINCETON UNIVERSITY AND NBER

REFERENCES

Akerlof, George, "The Market for Lemons: Qualitative Uncertainty and the Market Mechanism," *Quarterly Journal of Economics*, 84 (1970), 488–500.

Aumann, Robert, "Agreeing to Disagree," *Annals of Statistics*, 4 (1976), 1236–1239.

Barberis, Nicholas, and Richard Thaler, "A Survey of Behavioral Finance," in *Handbook of the Economics of Finance*, George Constantinides, Milton Harris, and Rene Stulz, eds. (Amsterdam: North-Holland, 2003).

Basak, Suleyman, "A Model of Dynamic Equilibrium Asset Pricing with Heterogeneous Beliefs and Extraneous Risk," *Journal of Economic Dynamics and Control*, 24 (2000), 63–95.

Bernheim, Douglas, and Antonio Rangel, "Toward Choice-theoretic Foundations for Behavioral Welfare Economics," *American Economic Review Papers and Proceedings*, 97 (2007), 464–470.

———, "Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics," *Quarterly Journal of Economics*, 124 (2009), 51–104.

Bewley, Truman, "Knightian Decision Theory, Part I," *Decisions in Economics and Finance*, 25 (2002), 79–110.

Bianchi, Javier, Emine Boz, and Enrique Mendoza, "Macro-prudential Policy in a Fisherian Model of Financial Innovation," University of Maryland working paper, 2012.

Black, Fischer, "Noise," *Journal of Finance*, 41 (1986), 529–543.

Bolton, Patrick, Jose Scheinkman, and Wei Xiong, "Executive Compensation and Short-termist Behavior in Speculative Markets," *Review of Economic Studies*, 73 (2006), 577–610.

Brunnermeier, Markus, Christian Gollier, and Jonathan A. Parker, "Optimal Beliefs, Asset Prices, and the Preference for Skewed Returns," *American Economic Review Papers and Proceedings*, 97 (2007), 159–165.

Brunnermeier, Markus, and Jonathan A. Parker, "Optimal Expectations," *American Economic Review*, 95 (2005), 1092–1118.

Buraschi, Andrea, and Alexei Jiltsov, "Model Uncertainty and Option Markets with Heterogeneous Beliefs," *Journal of Finance*, 61 (2006), 2841–2897.

Cao, Dan, "Collateral Shortages, Asset Price and Investment Volatility with Heterogeneous Beliefs," MIT working paper, 2010.

David, Alexander, "Heterogeneous Beliefs, Speculation, and the Equity Premium," *Journal of Finance*, 63 (2008), 41–83.

Davila, Eduardo, "Optimal Financial Transaction Taxes," Harvard University working paper, 2014.

DellaVigna, Stefano, "Psychology and Economics: Evidence from the Field," *Journal of Economic Literature*, 47 (2009), 315–372.

Detemple, Jerome, and Shashidhar Murthy, "Intertemporal Asset Pricing with Heterogeneous Beliefs," *Journal of Economic Theory*, 62 (1994), 294–320.

Doskeland, Trond, and Hans Hvide, "Do Individual Investors Have Asymmetric Information Based on Work Experience?" *Journal of Finance* 66 (2011), 1011–1041.

Dreze, Jacques, "Market Allocations under Uncertainty," *European Economic Review*, 2 (1970), 133–165.

Dumas, Bernard, Alexander Kurshev, and Raman Uppal, "Equilibrium Portfolio Strategies in the Presence of Sentiment Risk and Excess Volatility," *Journal of Finance*, 64 (2009), 579–629.

Dumas, Bernard, Karen Lewis, and Emilio Osambela, "Differences of Opinion and International Equity Markets," University of Pennsylvania working paper, 2011.

Fostel, Ana, and John Geanakoplos, "Leverage Cycles and the Anxious Economy," *American Economic Review*, 98 (2008), 1211–1244.

Geanakoplos, John, "Liquidity, Default, and Crashes: Endogenous Contracts in General Equilibrium," in *Advances in Economics and Econometrics: Theory and Applications*, Eighth World Conference, Mathias Dewatripont, Lars Hansen, and Stephen Turnovsky, eds. (Cambridge, UK: Cambridge University Press, 2003), vol. 2, 170–205.

Geanakoplos, John, "The Leverage Cycle," in *NBER Macroeconomics Annual 2009*, Daron Acemoglu, Kenneth Rogoff, and Michael Woodford, eds. (Chicago: University of Chicago Press, 2010), 1–65.

Gennaioli, Nicola, Andrei Shleifer, and Robert Vishny, "Neglected Risks, Financial Innovation, Financial Fragility," *Journal of Financial Economics*, 104 (2012), 452–468.

Gilboa, Itzhak, Dov Samet, and David Schmeidler, "Utilitarian Aggregation of Beliefs and Tastes," *Journal of Political Economy*, 112 (2004), 932–938.

Gilboa, Itzhak, Larry Samuelson, and David Schmeidler, "No-Betting Pareto Dominance," *Econometrica*, 82 (2014), 1405–1442.

Gilchrist, Simon, Charles Himmelberg, and Gur Huberman, "Do Stock Price Bubbles Influence Corporate Investment?," *Journal of Monetary Economics*, 52 (2005), 805–827.

Grossman, Sanford, and Joseph Stiglitz, "On the Impossibility of Informationally Efficient Markets," *American Economic Review*, 70 (1980), 393–408.

Hammond, Peter, "Ex-ante and Ex-post Welfare Optimality under Uncertainty," *Economica*, 48 (1981), 235–250.

Harris, Richard, "Ex-post Efficiency and Resource Allocation under Uncertainty," *Review of Economic Studies*, 45 (1978), 427–436.

Harrison, Michael, and David Kreps, "Speculative Investor Behavior in a Stock Market with Heterogeneous Expectations," *Quarterly Journal of Economics*, 92 (1978), 323–336.

Hassan, Tarek, and Thomas Mertens, "The Social Cost of Near-Rational Investment," University of Chicago working paper, 2011.

He, Zhiguo, and Wei Xiong, "Debt Financing in Asset Markets," *American Economic Review Papers and Proceedings*, 102 (2012), 88–94.

Hirshleifer, David, "Investor Psychology and Asset Pricing," *Journal of Finance*, 56 (2001), 1533–1597.

Hong, Harrison, Jose Scheinkman, and Wei Xiong, "Asset Float and Speculative Bubbles," *Journal of Finance*, 61 (2006), 1073–1117.

Hong, Harrison, and David Sraer, "Quiet Bubbles," *Journal of Financial Economics*, 110 (2013), 596–606.

Jouini, Elyes, and Clotilde Napp, "Consensus Consumer and Intertemporal Asset Pricing with Heterogeneous Beliefs," *Review of Economic Studies*, 74 (2007), 1149–1174.

Koszegi, Botond, and Matthew Rabin, "Mistakes in Choice-Based Welfare Analysis," *American Economic Review*, 97 (2007), 477–481.

Kubler, F., and K. Schmedders, "Financial Innovation and Asset Price Volatility," *American Economic Review: Papers and Proceedings*, 102 (2012), 147–151.

Kreps, David, *Microeconomic Foundations I: Choice and Competitive Markets* (Princeton, NJ: Princeton University Press, 2012).

Kurz, Mordecai, "Rational Beliefs and Endogenous Uncertainty," *Economic Theory*, 8 (1996), 383–397.

Mas-Colell Andreu, Michael Whinston, and Jerry Green, *Microeconomic Theory* (Oxford: Oxford University Press, 1995).

Milgrom, Paul, and Nancy Stokey, "Information, Trade and Common Knowledge," *Journal of Economic Theory*, 26 (1982), 17–27.

Miller, Edward, "Risk, Uncertainty, and Divergence of Opinion," *Journal of Finance*, 32 (1977), 1151–1168.

Mongin, Philippe, "Spurious Unanimity and the Pareto Principle," Universite Cergy-Pontoise working paper, 1997.

Morris, Stephen, "Trade with Heterogeneous Prior Beliefs and Asymmetric Information," *Econometrica*, 62 (1994), 1327–1347.

———, "The Common Prior Assumption in Economic Theory," *Economics and Philosophy*, 11 (1995), 227–253.

———, "Speculative Investor Behavior and Learning," *Quarterly Journal of Economics*, 110 (1996), 1111–1133.

Panageas, Stavros, "The Neoclassical q Theory of Investment in Speculative Markets," University of Chicago working paper, 2006.

Posner, Eric, and Glen Weyl, "An FDA for Financial Innovation: Applying the Insurable Interest Doctrine to 21st Century Financial Markets," *Northwestern University Law Review*, 107 (2013), 1307–1485.

Sandroni, Alvaro, and Francesco Squintani, "Overconfidence, Insurance, and Paternalism," *American Economic Review*, 97 (2007), 1994–2004.

Savage, Leonard, *The Foundation of Statistics* (Hoboken, NJ: Wiley, 1954).

Scheinkman, Jose, and Wei Xiong, "Overconfidence and Speculative Bubbles," *Journal of Political Economy*, 111 (2003), 1183–1219.

Sebenius, James, and John Geanakoplos, "Don't Bet on It: Contingent Agreements with Asymmetric Information," *Journal of the American Statistical Association*, 78 (1983), 424–426.

Shen, Ji, Hongjun Yan, and Jinfan Zhang, "Collateral-Motivated Financial Innovation," Yale University working paper, 2011.

Sims, Christopher, "Inflation Expectations, Uncertainty, and Monetary Policy," Princeton University working paper, 2008.

Simsek, Alp, "Speculation and Risk Sharing with New Financial Assets," *Quarterly Journal of Economics*, 128 (2013a), 1365–1396.

———, "Belief Disagreements and Collateral Constraints," *Econometrica*, 81 (2013b), 1–53.

Spinnewijn, Johannes, "Unemployed but Optimistic: Optimal Insurance Design with Biased Beliefs," *Journal of the European Economic Association,* (2014); doi: 10.1111/jeea.12099.

Starr, Ross, "Optimal Production and Allocation under Uncertainty," *Quarterly Journal of Economics*, 87 (1973), 81–93.

Stiglitz, Joseph, "Using Tax Policy to Curb Speculative Trading," *Journal of Financial Services Research*, 3 (1989), 101–115.

Summers, Lawrence, and Victoria Summers, "When Financial Markets Work too Well: A Cautious Case for a Securities Transaction Tax," *Journal of Financial Services Research*, 3 (1989), 261–286.

Tirole, Jean, "Overcoming Adverse Selection: How Public Intervention can Restore Market Functioning," *American Economic Review*, 102 (2012), 29–59.

von Weizsäcker, Christian, "General Equilibrium and Uncertainty," (MIT teaching note, 1969).

Weyl, Glen, "Is Arbitrage Socially Beneficial?" Princeton University working paper 2007.

Williams, Allan, and Adrian Lund, "Seat Belt Use Laws and Occupant Crash Protection in the United States," *American Journal of Public Health*, 76 (1986), 1438–1442.

Wu, Ho-Mou, and Wen-Chung Guo, "Asset Price Volatility and Trading Volume with Rational Beliefs," *Economic Theory*, 23 (2004), 795–829.

Xiong, Wei, and Hongjun Yan, "Heterogeneous Expectations and Bond Markets," *Review of Financial Studies*, 23 (2010), 1433–1466.

Zapatero, Fernando, "Effects of Financial Innovations on Market Volatility when Beliefs Are Heterogeneous," *Journal of Economic Dynamics and Control*, 22 (1998), 597–626.

Zhuk, Sergei, "Bubbles and Technological Innovations," Princeton University working paper, 2012.

This page intentionally left blank