Groupthink:
Collective Delusions in Organizations and Markets

Roland Bénabou*

Princeton University

This version: December 2011

Abstract

This paper investigates collective denial and willful blindness in groups, organizations and markets. Agents with anticipatory preferences, linked through an interaction structure, choose how to interpret and recall public signals about future prospects. Wishful thinking (denial of bad news) is shown to be contagious when it is harmful to others, and self-limiting when it is beneficial. Similarly, with Kreps-Porteus preferences, willful blindness (information avoidance) spreads when it increases the risks borne by others. This general mechanism can generate multiple social cognitions of reality, and in hierarchies it implies that realism and delusion will trickle down from the leaders. The welfare analysis differentiates group morale from groupthink and identifies a fundamental tension in organizations' attitudes toward dissent. Contagious exuberance can also seize asset markets, generating investment frenzies and crashes.

*I am grateful for valuable comments to Daron Acemoglu, George Akerlof, Bruno Biais, Alan Blinder, Patrick Bolton, Philip Bond, Markus Brunnermeier, Andrew Caplin, Sylain Chassang, Rafael Di Tella, Xavier Gabaix, Bob Gibbons, Boyan Jovanovic, Alessandro Lizzeri, Glenn Loury, Kiminori Matsuyama, Wolfgang Pesendorfer, Ben Polak, Eric Rasmussen, Ricardo Reis, Jean-Charles Rochet, Tom Romer, Julio Rotemberg, Tom Sargent, Hyun Shin, David Sraer, Jean Tirole, Glen Weyl, Muhamet Yildiz and participants at many seminars and conferences. Rainer Schwabe, Andrei Rachkov and Edoardo Grillo provided superb research assistance. Support from the Canadian Institute for Advanced Research and the Institute for Advanced Study in Toulouse are gratefully acknowledged.

"The Columbia accident is an unfortunate illustration of how NASA's strong cultural bias and its optimistic organizational thinking undermined effective decision-making." (Columbia Accident Investigation Board Final Report, 2003)

"The ability of governments and investors to delude themselves, giving rise to periodic bouts of euphoria that usually end in tears, seems to have remained a constant. (Reinhart and Rogoff, "This Time Is Different: Eight Centuries of Financial Folly", 2009).

1. Introduction

In the aftermath of corporate and public-sector disasters, it often emerges that participants fell prey to a collective form of willful blindness and overconfidence: mounting warning signals were systematically cast aside or met with denial, evidence avoided or selectively reinterpreted, dissenters shunned. Market bubbles and manias exhibit the same pattern of investors acting "color-blind in a sea of red flags", followed by a crash.[1] To shed light on these phenomena, this paper analyzes how distorted beliefs spread through organizations such as firms, bureaucracies and markets.

Janis (1972), studying policy decisions such as the Bay of Pigs invasion, the Cuban missile crisis and the escalation of the Vietnam war, identified in those that ended disastrously a cluster of symptoms for which he coined the term "groupthink".[2] Although later work was critical of his characterization of those episodes, the concept has flourished and spurred a large literature in social and organizational psychology. Defined in Merriam-Webster's dictionary as *"a pattern of thought characterized by self-deception, forced manufacture of consent, and conformity to group values and ethics",* groupthink was strikingly documented in the official inquiries conducted on the Challenger and Columbia space shuttle disasters. It has also been invoked as a contributing factor in the failures of companies such as Enron and Worldcom, decisions relating to the second Iraq war, and the recent financial crisis.[3] At

[1] I borrow here the evocative title of Norris' (2008) account of Merrill Lynch's mortgage securitization debacle. A year later, the Inspector General's Report (2009) on the SEC's failure concerning the Madoff scheme contained over 130 mentions of "red flags".

[2] The eight symptoms were: (a) illusion of invulnerability; (b) collective rationalization; (c) belief in inherent morality; (d) stereotyped views of out-groups; (e) direct pressure on dissenters; (f) self-censorship; (g) illusion of unanimity; (h) self-appointed mindguards. The model developed here will address (a) to (g).

[3] On the shuttle accidents, see Rogers Commission (1986) and Columbia Accident Investigation Board (2003). On Enron, see Samuelson (2001), Cohan (2002), Eichenwald (2005) and Pearlstein (2006). On Iraq, see e.g., Hersh (2004), Suskind (2004) and Isikoff and Corn (2007).

the same time, one must keep in mind that the mirror opposite of harmful "groupthink" is valuable "group morale" and therefore ask how the two mechanisms differ, even though both involve the maintenance of collective optimism despite negative signals.

To analyze these issues, I develop a model of (individually rational) *collective denial and willful blindness.* Agents are engaged in a joint enterprise where their final payoff will be determined by their own action and those of others, all affected by a common productivity shock. To distinguish groupthink from standard mechanisms, there are no complementarities in payoffs, nor any private signals that could give rise to herding or social learning. Each agent derives anticipatory utility from his future prospects, and consequently faces a tradeoff: he can accept the grim implications of negative public signals about the project's value (realism) and act accordingly, or maintain hopeful beliefs by discounting, ignoring or forgetting such data (denial), at the risk of making overoptimistic decisions.

The key observation is that this tradeoff is shaped by how others deal with bad news, creating cognitive linkages. When an agent benefits from others' overoptimism, his improved prospects make him more accepting of the bad news which they ignore. Conversely, when he is made worse off by others' blindness to adverse signals, the increased loss attached to such news pushes him toward denial, which is then contagious. Thinking styles thus become strategic substitutes or complements, depending on the sign of externalities (not cross-partials) in the interaction payoffs. When interdependence among participants is high enough, this *Mutually Assured Delusion* (MAD) principle can give rise to multiple equilibria with different "social cognitions" of the same reality. The same principle also implies that, in organizations where some agents have a greater impact on others' welfare than the reverse (e.g., managers on workers), strategies of realism or denial will "trickle down" the hierarchy, so that subordinates will in effect *take their beliefs from the leader.*

The underlying insight is quite general and, in particular, does not depend on the assumptions of anticipatory utility and malleable memory or awareness. To demonstrate this point, I analyze a variant of the model in which both are replaced by Kreps-Porteus (1978) preferences for late resolution of uncertainty. This also serves, importantly, to address collective willful ignorance (ex-ante avoidance of information) in the same way as the benchmark model addresses collective denial (ex-post distortion of beliefs). In line with the MAD principle, I show that if an agent's remaining uninformed about the state of the world leads him to

increase the *risks* born by others, this pushes them toward also delaying becoming informed; as a result, ignorance becomes contagious and risk spreads through the organization. Conversely, when information avoidance has beneficial hedging spillovers, it is self-dampening.[4]

The model's welfare analysis makes clear what factors distinguish valuable group morale from harmful groupthink, irrespective of anticipatory payoffs, which average out across states of the world. It furthermore explains why organizations and societies find it desirable to set up ex-ante commitment mechanisms protecting and encouraging dissent (constitutional guarantees of free speech, whistle-blower protections, devil's advocates, etc.), even when ex-post everyone would unanimously want to ignore or "kill" the messengers of bad news.

In market interactions, finally, prices typically introduce a substitutability between supply decisions that works against collective belief. Nonetheless, in asset markets with limited liquidity (new types of securities, startup firms, housing), *contagious exuberance* can again take hold**,** leading to investment frenzies followed by deep crashes. When signals about fundamentals turn from green to red, each participant who keeps investing contributes to driving the final market-clearing price further down. This makes it ultimately more costly for others to also overinvest, but at the same time magnifies the capital losses that realism would require them to immediately acknowledge on their outstanding positions. In equilibrium the stock effect dominates the flow effect, so that all prefer to keep believing in strong fundamentals than recognize the warning signals of a looming crash.

1.1. Related evidence

Asymmetric updating and information avoidance. Besides the vast literature on overconfidence and overoptimism, there is a long-standing body of work more specifically documenting people's tendency to selectively process, interpret and recall data in ways that lead to more favorable beliefs about their own traits or future prospects.[5] While earlier studies relied on self-reports rather than incentivized choices, several recent papers offer rigorous confirma-

[4]Thus, as in the benchmark (anticipatory utility) version, agents' "patterns of thought" become substitutes or complements in a way that turns entirely on the first derivatives of the payoff structure. The difference is that these externalities now operate on the variance rather than the conditional expectation of agents' utilities. The MAD mechanism is also shown to be robust along many other dimensions, such as nonseparable payoffs or limited sophistication (adaptive learning).

[5]See, e.g., Mischel et al. [1976] and Thompson et al. [1992] on the differential recall of favorable and unfavorable, information, and Kunda [1987] on the biased processing of self-relevant data.

tions of a *differential response to good and bad news.* Eil and Rao (2010) and Möbius et al. (2010) provide subjects with several rounds of objective data on their IQ rankings; the first paper uses physical attractiveness as well. They also elicit, using incentive-compatible scoring rules, subjects' prior and posterior beliefs about their rank. Eil and Rao find that, compared with Bayes' rule, subjects systematically underrespond to negative news and are much closer to proper updating for positive news. Möbius et al. similarly find significant underupdating in response to bad news; subjects also update less than fully in response to good news, but the gap with Bayes' rule is significantly smaller. In both studies, a significant fraction of subjects also display *information aversion*, paying money to avoid learning their exact IQ or beauty score after the last round.[6]

Mijovic-Prelec and Prelec (2010) demonstrate costly self-deception about the likelihood of an exogenous binary event: although incentivized for accuracy, subjects reverse their predictions as a function of their stakes in the outcome. Similarly, Mayraz (2011) finds that subjects assigned to be buyers or sellers at some future price make (incentivized) predictions about it that vary systematically with their monetary stakes in its being high or low. These results establish the role of the anticipatory motive in belief distortion and show that the latter responds to incentives, as will be the case in the model. Hedden et al. (2008) use FMRI on subjects engaged in the first paper's task to identify the neural correlates of self-deception. Self-deceivers (as revealed by their more systematic prediction reversals) exhibit distinctive activity patterns in regions of the brain generally associated to reward processing and in those associated with attentional and cognitive control.

In the field, Choi and Lou (2010) find evidence of self-serving, asymmetric updating by mutual fund managers. Using a large panel of actively managed funds, they measure a manager's confidence in his stock-picking ability or private signal quality by the deviation, attributable to his active trades, between his portfolio weights and the relevant market index. Following confirming signals (positive realized excess returns over the year), fund managers trade more actively, thereby exhibiting increased self-confidence. Following dis-

[6] In contrast, no updating bias or information avoidance occurs when rank is randomly assigned. For self-relevant information, both findings of underadjustment to bad news and a lesser underadjustment (possibly none) to good news accord well with the awareness-management model of Bénabou and Tirole (2002), which corresponds to equation (6) below (see also footnote 26). The experimental design is also such that the feedback received by participants is not credibly transferable to outsiders (and, for beauty, redundant), ruling out any Hirshleifer-type effect to explain a negative value of information.

confirming ones (negative realized excess returns) there is no equivalent decrease –in fact, zero adjustment cannot be rejected. Furthermore, this selective updating leads to suboptimal investments, as positive past excess returns are found to negatively predict subsequent risk-adjusted fund performance. Individual investors also display a good-news / bad news asymmetry, both in the recall of their portfolios' past returns (Goetzman and Peles (1997)) and in informational decisions, where far more go online to look up the value of their portfolios on days when the market is up than when it is down (Karlsson et al. (2009)).

The avoidance of decision-relevant information for fear of learning of a bad outcome is extensively documented in the medical sphere, where significant fractions of people avoid checkups, refuse to take tests for HIV infection or genetic predispositions to certain cancers, even when anonymity is ensured and in countries with universal health insurance and strict anti-discrimination regulations. This body of evidence and its relationship to anticipatory anxiety are reviewed in Caplin and Leahy (2001) and Caplin and Eliaz (2003).

Organizational and market blindness. These individual propensities to cognitive distortion naturally raise the question of equilibrium: what environments will make such behaviors socially contagious or self-limiting, and with what welfare implications? Surprisingly, this question has never been considered, even in the large literature on informational attitudes that followed Kreps and Porteus (1978). Yet the issue is not only theoretically interesting, but also potentially important to make sense of notions such as "optimistic organizational thinking" and "governments and investors deluding themselves".

While there is yet no formal study of motivated cognition at the level of a firm or market, a number of case studies and official investigation reports provide supporting evidence for the idea.[7] I summarize in online Appendix D several "patterns of denial" –including, once again, actively avoiding information ex-ante and changing standards of evidence ex-post– that recur strikingly from NASA to the FED, SEC and Fannie Mae, from Enron to investment banks, AIG and individual investors.[8] The historical studies of financial crises by Mckay (1980),

[7] This type of data is of course largely qualitative and selected on failure, but also notable for its depth and extensiveness: records of meetings, confidential emails and memos, sworn testimony, financial transactions, technical tests and analyses by experts.

[8] Another point made there is the insufficiency of pure moral hazard as the sole explanation. Instead, self-serving rationalizations ("ethical fading", e.g., Tenbrunsel, and Messick (2004), Bazerman and Tenbrunsel (2011)) and overoptimistic hubris are key enablers of most corporate misconduct and financial fraud (see also Huseman and Driver (1979), Sims (1992), Anand et al. (2005) and Schrand and Zechman (2008)).

Kindleberger and Aliber (2005), Shiller (2005) and Reinhart and Rogoff (2009) provide many similar examples, from which their conclusions of contagious "delusions", "manias", "irrational exuberance" and "financial folly" are derived.

1.2. Related theories

This work ties into multiple literatures. The first one centers on cognitive dissonance and other forms of self-deception, the second on anticipatory feelings and attitudes toward information.[9] Most papers so far have focused on individual rather than social beliefs, and none has asked what makes wishful thinking infectious or self-limiting. The analysis of group morale and groupthink in organizations relates the paper to a third line of work, which deals with heterogeneous beliefs and overoptimism in firms.[10] Beliefs there are most often exogenous (reflecting different priors), whereas here they endogenously spread, horizontally or vertically, through all or part of the organization. Beyond economics, the paper relates to the work in management on corporate culture and to that in psychology on "social cognition".

In models of social conformity and in models of herding, collective errors arise from divergences between individuals' private signals and their publicly observable statements or actions. Departing from these standard channels, the paper identifies a novel mechanism generating interdependent beliefs and behaviors, which: (i) requires neither private information nor lack of anonymity; (ii) accounts for both conformism and contrarianism, with clear predictions as to when each should be observed; (iii) is in line with the micro-experimental and case-study evidence of biased updating and information avoidance; (iv) generates many distinctive and potentially testable comparative-statics results.

A first alternative source of group error is social pressure to conform.[11] For instance, if

[9] On cognitive dissonance, see Akerlof and Dickens (1982), Schelling (1986), Kuran (1993), Rabin (1994), Bénabou and Tirole (2002, 2004, 2006b), Compte and Postlewaite (2004) and Di Tella et al. (2007). On anticipation, see Loewenstein (1987), Caplin and Leahy (2001), Landier (2000), Caplin and Eliaz (2005), Brunnermeier and Parker (2005), Bernheim and Thomadsen (2005), Köszegi (2006, 2010), Eliaz and Spiegler (2006), Brunnermeier et al. (2007) and Bénabou and Tirole (2011). For an evolutionary account of self-deception see, e.g., von Hippel and Trivers (2011), who argue that it initially evolved to facilitate the deception of others, but once developed also affected different aspects of behavior.

[10] On the theoretical side, see, e.g. Rotemberg and Saloner (1993), Bénabou and Tirole (2003), Fang and Moscarini (2005), Van den Steen (2005), Gervais and Goldstein (2007) and Landier et al. (2009). On the empirical side, see, e.g., Malmendier and Tate (2005, 2008) or Camerer and Malmendier (2007).

[11] One could also invoke some exogenous preference for agreeing with the majority, but this has little predictive content as to which settings are more conducive to the phenomenon (e.g., congruent or dissimilar objectives), or whether conformist preferences apply to genuine beliefs or only publicly stated opinions.

agents are heard or seen by both a powerful principal (boss, group leader, government) and third parties whom he wants to influence, they may just toe the line for fear of retaliation. Their true beliefs should still show up ex-post in any unmonitored actions they were able to take, yet in many cases of organizational failure no such discrepancy is observed.[12] Self-censorship should also not occur when agents can communicate separately with the boss, who should then want to hear both good and bad news. There are nonetheless many instances where deliberately confidential and highly credible warnings were flatly ignored, with disastrous consequences for the decision-maker.[13]

A second important source of conformity is signaling or career concerns. Thus, when the quality of their information is unknown, agents whose opinion is at odds with most already expressed may keep it to themselves, for fear of appearing incompetent or lazy (Ottaviani and Sørensen (2001), Prat (2005)). Significant mistakes in group decisions can result, in contexts where differential information is important and anonymous communication or voting not feasible.[14] The mechanism explored here, by contrast, is portable between environments with and without anonymity, including financial markets and the electoral arena, where investors and voters make decisions privately.

The model's application to market manias and crashes links the paper to the literatures on bubbles and herding, but the mechanism is very different from those of existing models. First, in a standard cascade, each investor acts exactly as a cool-headed and benevolent statistician would advise him to. He thus goes against his own signal only in instances where the herd is truly more likely to have it right, and more generally displays the usual desire for accurate knowledge.[15] This seems a far cry from the wishful assumptions and

Relatedly, the search for robust comparative-statics regularities in Asch-like (1956) conformity experiments has been largely unsuccessful (Lalancette and Standing (1990)).

[12] Thus, in the weeks and days preceding the collapses of Enron and Lehman Brothers, most employees did not significantly alter or hedge their retirement portfolios, which were heavily loaded with company stock.

[13] For instance, Enron V.P. Sharon Watkins' memo to CEO Ken Lay, and FED governor Edward Gramlich's warnings to Chairman Greenspan (see online Appendix D).

[14] With anonymity, information aggregation should be achievable even when agents have different a priori levels of expertise, by allocating ballots in proportion to these priors. Private information is also not always a key issue in collective errors. In the two space shuttle disasters, for instance, NASA mission managers and engineers were all looking at the same data; see online Appendix D.

[15] See, e.g., Banerjee (1992), Bikhchandani, Hirshleifer and Welch (1992), Caplin and Leahy (1994), Chamley and Gale (1994). In versions of herding models with naive agents (e.g., Eyster and Rabin (2009)), agents put excessive weight on the actions of others, but still without any kind of wishful thinking or motivated reasoning –they just lack statistical or strategic sophistication. Experimental tests show that people in fact overweigh their *own* information (a form of overconfidence) relative to that embodied in other players' moves,

rationalizations ("new economy", this "time is different", "they are not making any more land", etc.) repeatedly described by observers and historians. Second, in herding models the problem is a failure to aggregate private signals, which becomes less relevant when more of this data becomes common knowledge, for example through statistical sources or the media. In market groupthink, by contrast, investors have access to very similar information, but their processing of it is distorted by a contagious form of motivated thinking.[16]

Section 2 presents the benchmark model and propositions on collective realism and denial. Section 3 derives related results for risk preferences and the contagion of informational attitudes. Section 4 examines welfare and the treatment of dissent. Section 5 deals with asset-market manias and crashes, and Section 6 concludes. In online Appendix B, the model is extended to deal with fatalism and apathy in the face of crises. Key proofs are gathered in the paper's main Appendix A, more technical ones in online Appendix C.

2. Groupthink in teams and organizations

2.1. Benchmark model

• *Technology.* A group of risk-neutral agents, $i \in \{1, ...n\}$, are engaged in a joint project (team, firm, military unit) or other activities generating spillovers. At $t = 1$, each chooses effort $e^i = 0$ or 1, with cost $ce^i, c > 0$. At $t = 2$, he will reap expected utility

$$U_2^i \equiv \theta \left[\alpha e^i + (1-\alpha)e^{-i}\right], \tag{1}$$

where $e^{-i} \equiv \frac{1}{n-1}\sum_{j\neq i} e^j$ denotes the average effort of others and $1-\alpha \in [0, 1-1/n]$ the degree of interdependence, reflecting the joint nature of the enterprise or the presence of cross-interests.[17] Depending on α, the choice of e^i ranges from a pure private good (or bad) to a pure public one. This payoff structure is maximally simple: all agents play symmetric roles, there is a fixed value to inaction $e = 0$, normalized to 0, and *no interdependence of any*

making cascades relatively rare and short-lived (e.g., Goeree et al. (2007), Weiszacker (2010)).

[16] In the recent financial crisis, most of the key data on household debt, mounting mortgage defaults, historical boom and bust cycles in real estate prices, etc., was easily accessible to the major players, including regulators, and even loudly advertised by a few but prominent Cassandras.

[17] Another source of interdependence is social preferences: altruism, family or kinship ties, social identity, etc. Thus, (1) is equivalent to $U_2^i \equiv \beta\theta e^i + (1-\beta)U_2^{-i}$ with $1-\alpha \equiv (1-\beta)(n-1)/(n-\beta)$. Altruistic concerns are explicitly studied in online Appendix B.

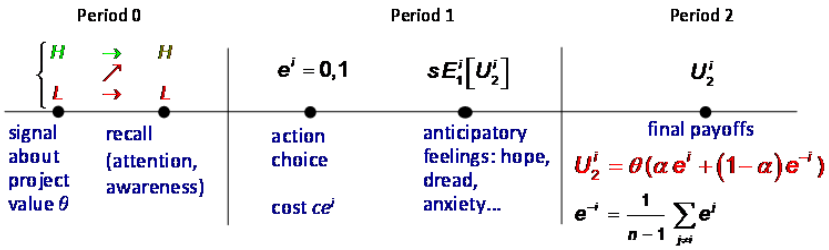

Figure 1: Timeline

kind between effort decisions. These assumptions serve only to highlight the key mechanism, and are all relaxed later on.

The productivity of the venture is a priori uncertain: its expected value is $\theta = \theta_H$ in state H and $\theta = \theta_L$ in state L, with $\Delta\theta \equiv \theta_H - \theta_L > 0$ and $\theta_H > 0$ without loss of generality. Depending on the context, θ can represent the value of a firm's product or business plan, the state of the market, the suitability of a political or military strategy, or the quality of a leader. Given (1), θ defines the expected social value of a choice $e^j = 1$, *relative* to what the alternative course of action would yield. Thus, for $\theta_L \geq 0$ each agent would prefer that others always choose $e^j = 1$, whereas for $\theta_L < 0$ he would like them to pursue the "appropriate" course of action for the organization, choosing $e^j = 1$ in state H and $e^j = 0$ in state L.[18]

• *Preferences.* The payoffs received during period 1 include the cost of effort, $-ce^i$, but also the *anticipatory utility* experienced from thinking about one's future prospects, $sE_1^i[U_2^i]$, where $s \geq 0$ (for "savoring" or "susceptibility") parametrizes the well-documented psychological and health effects of hopefulness, dread and similar emotions.[19]

At the start of period 1, agent i chooses effort to maximize the expected present value of payoffs, discounted at rate $\delta \in (0, 1]$:

$$U_1^i = -ce^i + sE_1^i\left[U_2^i\right] + \delta E_1^i\left[U_2^i\right]. \tag{2}$$

Given (1), his effort is determined solely by his beliefs about θ : $e^i = 1$ if $(s + \delta)\alpha E_1^i[\theta] > c$,

[18] It is thus not the sign of θ_L per se that is relevant, but how θ_L compares to the (social) return to taking the alternative action $e = 0$ in state L. The latter's normalization to zero is relaxed in Section 2.4.

[19] The parameter s typically increases with the length of period 1, during which uncertainty remains. The linear specification $sE_1^i[U_2^i]$ avoids building in either information-loving or information aversion (which will be studied in Section 3).

independently of what any one else may be doing. I shall assume that

$$\theta_L < \frac{c}{(s+\delta)\alpha} < \frac{c}{\delta\alpha} < q\theta_H + (1-q)\theta_L. \tag{3}$$

An agent acting on his sole prior will thus choose $e^i = 1$, whereas one who knows for sure that the state is L will abstain. Actual beliefs at $t = 1$ will depend on the news received at $t = 0$ and how objectively or subjectively the agent processes them, as described below. In doing so, he aims to maximize the discounted utility of all payoffs

$$U_0^i = -M^i + \delta E_0^i \left[-ce^i + sE_1^i \left[U_2^i\right]\right] + \delta^2 E_0^i \left[U_2^i\right], \tag{4}$$

where E_t^i denotes expectations at $t = 0, 1$ and M^i the date-0 costs of his cognitive strategy.

The main behavioral implications of these preferences arise from the tradeoff between accurate and hopeful beliefs embodied in (4).[20] To the extent that his cognitive "technology" allows it, an agent will update in a distorted manner (underadjusting to bad news as in Rao and Eil (2010) and Möbius et al. (2010)), and consequently invest even after seeing data showing that he should not. In short, he will engage in *wishful thinking.*[21]

• *Information and beliefs.* To represent agents' "patterns of thought", I use an extended version of the selective-recall technology in Bénabou and Tirole (2002). At $t = 0$, everyone observes a common signal that defines the state of the world: $\sigma = H, L$, with probabilities q and $1 - q$ respectively.[22] Each agent then chooses (consciously or not) how much attention to pay to the news, how to interpret it, whether to "keep it in mind" or "not think about it", etc. Formally, after observing σ he can:

(a) Accept the facts realistically, truthfully encoding $\hat{\sigma}^i = \sigma$ into memory or awareness (his date-1 information set).

(b) Engage in denial, censoring or rationalization, encoding $\hat{\sigma}^i = H$ instead of $\sigma = L$, or $\hat{\sigma}^i = L$ instead of $\sigma = H$. In addition to impacting later decisions, this may entail an

[20] On the general behavioral implications of models with utility from anticipation, see Köszegi (2010).

[21] Namely, "the attribution of reality to what one wishes to be true or the tenuous justification of what one wants to believe" (Merriam Webster), and "the formation of beliefs and making decisions according to what might be pleasing to imagine instead of by appealing to evidence, rationality or reality" (Wikipedia).

[22] Note that θ_σ is only the *expected* value of the project conditional on σ, so a low (high) signal need not preclude a high (low) final realization.

immediate cost $m \geq 0$.[23]

(c) When indifferent between these two courses of actions, use a mixed strategy.[24]

This simple informational structure captures a broad range of situations. The perfect correlation between agents' signals could be relaxed, but serves to make clear that the model has nothing to do with herding or cascades, where privately informed agents make inferences from each other's behavior. The prior distribution $(q, 1-q)$ could be conditional on some earlier signal being good news, such as the appearance of a new technology or market opportunity. These positive news may also have warranted some initial investment in the activity, including the formation of the group itself.

Intuition suggests that it is only in state L that an agent may censor his signal: given (1) and the utility from anticipation, he would never want to substitute bad news for good ones.[25] Verifying in Appendix C that such is indeed the case as long as $m > 0$, no matter how small, I focus here on cognitive decisions in state L and denote

$$\lambda^i \equiv \Pr\left[\hat{\sigma}^i = L | \sigma = L\right] \tag{5}$$

the awareness strategy of agent i. Later on I will consider payoffs structures more general than (1), under which either state may be censored.

While people can selectively process information, their latitude to self-deceive is generally not unconstrained. At $t = 1$, agent i no longer has direct access to the original signal, but if he is aware of his tendency to discount bad news he will take it into account. Thus, when

[23] This can involve material resources (eliminating evidence, avoiding certain people, searching for and rehearsing desirable signals) or mental ones (stress from repression, cognitive dissonance, guilt). As explained below, any arbitrarily small $m > 0$ suffices to rule out uninteresting equilibria in which there is signal distortion in both states ("inefficient encoding"). Beyond this, all the paper's key results apply equally with $m = 0$, though non-zero costs are more realistic, particularly for the welfare analysis.

[24] Agents thus do not commit in advance to a (state-contingent) mixture of realism and denial, but respond optimally to the news they receive. It seems unlikely that someone could constrain a priori how he will interpret or recall different signals, particularly in a social context where he may be exposed to others' response to the news. Such commitment is more plausible at the organizational level, and this is analyzed in Section 4. For a sophisticated Bayesian, cognitive commitment (when feasible) would be equivalent to coarsening the signal structure $\sigma = H, L$; such ex-ante informational choices are studied in Section 3.

[25] An agent who likes pleasant surprises and dislikes disappointments, on the other hand, may want to. Such preferences correspond (maintaining linearity) to $s = -\delta s'$, $0 < s' < 1$, so that the last two terms in (4) become $\delta^2 E_0^i \left[U_2^i - s' E_1^i \left[U_2^i\right]\right]$. All the results could be transposed to the case $s < 0$, leading to a (less empirically relevant) model of collective "defensive pessimism". By focussing on $s \geq 0$, I am implicitly assuming that the disappointment-aversion motive, if present, is dominated by anticipatory concerns. Such is the case, for instance, if the"waiting" period 1 is long enough. The potential social or evolutionary value of anticipatory concerns is discussed in Section 4.

$\hat{\sigma}^i = L$ he knows for sure that the state is L, but when $\hat{\sigma}^i = H$ his posterior belief is only

$$\Pr\left[\sigma = H \mid \hat{\sigma}^i = H, \lambda^i\right] = \frac{q}{q + (1-q)\chi(1-\lambda^i)} \equiv r(\lambda^i), \tag{6}$$

where λ^i is his equilibrium rate of realism (awareness of bad news) and $\chi \in [0,1]$ parametrizes cognitive sophistication. I shall focus on the benchmark case of rational Bayesians ($\chi = 1$), but the analysis goes through for any χ, including full naiveté ($\chi = 0$).[26]

To analyze the equilibria of this game, I proceed in three steps. First, I fix everyone but agent i's awareness strategy at some arbitrary $\lambda^{-i} \in [0,1]$ and look for his "best response" λ^i.[27] Second, I identify the general principle that governs whether individual cognitions are strategic *substitutes* (the more others delude themselves, the better informed I want to be) or *complements* (the more others delude themselves, the less I also want to face the truth). Finally, I derive conditions under which groupthink arises in its most striking form, where both collective realism and collective denial constitute self-sustaining *social cognitions.*

2.2. Best-response awareness

Following bad news, agents who remain aware that $\theta = \theta_L$ do not exert effort, while those who managed to ignore or rationalize away the signal have posterior $r(\lambda^j) \geq q$ and choose $e^j = 1$. Responding as a realist to a signal $\sigma = L$ thus leads for agent i to intertemporal expected utility (R is for "realism")

$$U^i_{0,R} = \delta(\delta + s)\left[\alpha \cdot 0 + (1-\alpha)(1-\lambda^{-i})\right]\theta_L, \tag{7}$$

reflecting his knowledge that only the fraction $1 - \lambda^{-i}$ of other agents who are in denial will exert effort. If he censors, on the other hand, he will assign probabilities $r(\lambda^i)$ to the state being H, in which case everyone exerts effort with productivity θ_H, and $1 - r(\lambda^i)$ to it being

[26] The paper's positive results become only stronger with $\chi < 1$, as self-deception is more effective. In the welfare analysis, an extra term is simply added to the criterion computed with $\chi = 1$; see footnote 50. Note also that (6) generates both empirical findings discussed in footnote 6, for any $\lambda^i < 1$ and $\chi < q/(1-q)$.

[27] With imperfect recall, each agent's problem is itself a game of strategic information transmission between his date-0 and date-1 "selves". Condition (3) and $m > 0$ will rule out any multiplicity of intrapersonal equilibria, simplifying the analysis and making clear that the groupthink phenomenon is one of *collectively sustained* cognitions. Note also that the focus on symmetric group equilibria, implicit in equating all λ^j's to a common λ^{-i}, is without loss of generality when there are many identical agents, as all best-respond to the aggregate. For finite n and/or heterogenous groups, there can also be asymmetric equilibria; see Section 2.4.

really L, in which case only the other optimists like him are working and their output is $(1-\lambda^{-i})\theta_L$. Hence (D is for "denial"):

$$U^i_{0,D} = -m+\delta\left(-c+\delta\left[\alpha+(1-\alpha)(1-\lambda^{-i})\right]\theta_L\right) \\ +\delta s\left(r(\lambda^i)\theta_H+\left(1-r(\lambda^i)\right)\left[\alpha+(1-\alpha)(1-\lambda^{-i})\right]\theta_L\right). \tag{8}$$

Agent i's incentive to deny reality, given that a fraction $1-\lambda^{-i}$ of others do so, is thus:

$$U^i_{0,D}-U^i_{0,R} = -m-\delta\left[c-(\delta+s)\,\alpha\theta_L\right]+\delta s r(\lambda^i)\left[(1-\alpha)\lambda^{-i}\theta_L+\Delta\theta\right]. \tag{9}$$

The second term is the net loss from mistakenly choosing $e^i=1$ due to overoptimistic beliefs**.** The third term is the gain in anticipatory utility, proportional to s and the posterior belief $r(\lambda^i)$ that the state is H, which has two effects. First, the agent raises his estimate of the fraction choosing $e=1$, from $1-\lambda^{-i}$ to 1; at the true productivity θ_L, this contributes $(1-\alpha)\lambda^{-i}\theta_L$ to his expected welfare. Second, he believes the project's value to be θ_H rather than θ_L, so that when everyone chooses $e=1$ his welfare is higher by $\Delta\theta=\theta_H-\theta_L$.

Let $\Psi(\lambda^i,s|\lambda^{-i})$ denote the right-hand side of (9), representing agent i's net incentive for denial. Since it is increasing in his "habitual" degree of realism λ^i, there is a unique fixed point (personal equilibrium), which characterizes the optimal awareness strategy:

(a) $\lambda^i=1$ if $\Psi(1,s|\lambda^{-i})\le 0$. By (9), and noting that $\alpha\theta_L+\Delta\theta+(1-\alpha)\lambda^{-i}\theta_L \ge \min\{\Delta\theta,\theta_H\}>0$, this means

$$s\le\frac{m/\delta+c-\delta\alpha\theta_L}{\alpha\theta_L+\Delta\theta+(1-\alpha)\lambda^{-i}\theta_L}\equiv\underline{s}(\lambda^{-i}). \tag{10}$$

(b) $\lambda^i=0$ if $\Psi(0,s|\lambda^{-i})\ge 0$. By (9), and noting that $\alpha\theta_L+q\left[\Delta\theta+(1-\alpha)\lambda^{-i}\theta_L\right]\ge \min\{q\Delta\theta,q\theta_H+(1-q)\,\theta_L\}>\min\{q\Delta\theta,c/(s+\delta)\}>0$, this means

$$s\ge\frac{m/\delta+c-\delta\alpha\theta_L}{\alpha\theta_L+q\left[\Delta\theta+(1-\alpha)\lambda^{-i}\theta_L\right]}\equiv\bar{s}(\lambda^{-i}). \tag{11}$$

Moreover, $\underline{s}(\lambda^{-i})<\bar{s}(\lambda^{-i})$, since $\Delta\theta+(1-\alpha)\lambda^{-i}\theta_L\ge\Delta\theta+(1-\alpha)\lambda^{-i}\min\{\theta_L,0\}\ge\Delta\theta+\min\{\theta_L,0\}=\min\{\theta_H,\Delta\theta\}>0$.

(c) $\lambda^i\in(0,1)$ is the unique solution to $\Psi(\lambda^i,s|\lambda^{-i})=0$ for $\Psi(0,s|\lambda^{-i})<0<\Psi(1,s|\lambda^{-i})$,

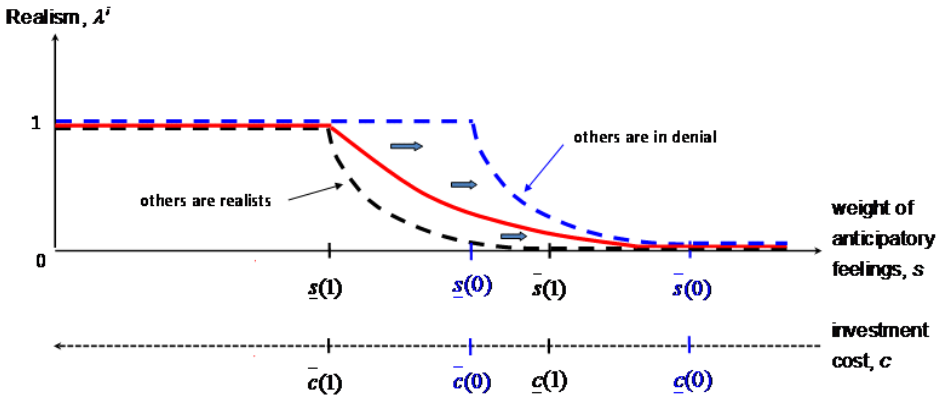

Figure 2: Group Morale ($\theta_L > 0$). The dashed lines give agent i's optimal awareness λ^i when others are realists ($\lambda^j = 1$) or deniers ($\lambda^j = 0$); arrows indicate the shift between the two. The solid line defines the social equilibrium.

which corresponds to $\underline{s}(\lambda^{-i}) < s < \bar{s}(\lambda^{-i})$.

This *best response to how others think* is illustrated by the dashed curves in Figures 2-3, as a function of either s or c, which have opposite effects. Variations in s provide more transparent intuitions (e.g., $s = 0$ is the classical benchmark), whereas variations in c are directly observable and experimentally manipulable. All results are therefore stated in a dual form that covers both approaches.

Lemma 1. ***(Optimal awareness)*** *For any cognitive strategy λ^{-i} used by other agents, there is a unique optimal awareness rate λ^i for agent i :*
(i) $\lambda^i = 1$ for s up to a lower threshold $\underline{s}(\lambda^{-i}) > 0$, λ^i is strictly decreasing in s between $\underline{s}(\lambda^{-i})$ and an upper threshold $\bar{s}(\lambda^{-i}) > \underline{s}(\lambda^{-i})$, and $\lambda^i = 0$ for s above $\bar{s}(\lambda^{-i})$.
(ii) Similarly, $\lambda^i = 0$ for c below a threshold $\underline{c}(\lambda^{-i})$, λ^i is strictly increasing in c between $\underline{c}(\lambda^{-i})$ and a threshold $\bar{c}(\lambda^{-i}) > \underline{c}(\lambda^{-i})$, and $\lambda^i = 0$ for c above $\underline{c}(\lambda^{-i})$.

As one would expect, the more important anticipatory feelings are to an agent's welfare, and the lower the cost of mistakes, the more bad news will be repressed. The next result brings to light the key insight concerning the *social* determinants of wishful thinking.

Proposition 1. ***(MAD principle)*** *(i) An agent's degree of realism λ^i decreases with that of others, λ^{-i}, (substitutability) if $\theta_L > 0$, and increases with it (complementarity) if $\theta_L < 0$.*
(ii) λ^i increases with the degree of spillovers $1 - \alpha$ if $\theta_L > 0$, and decreases if $\theta_L < 0$.

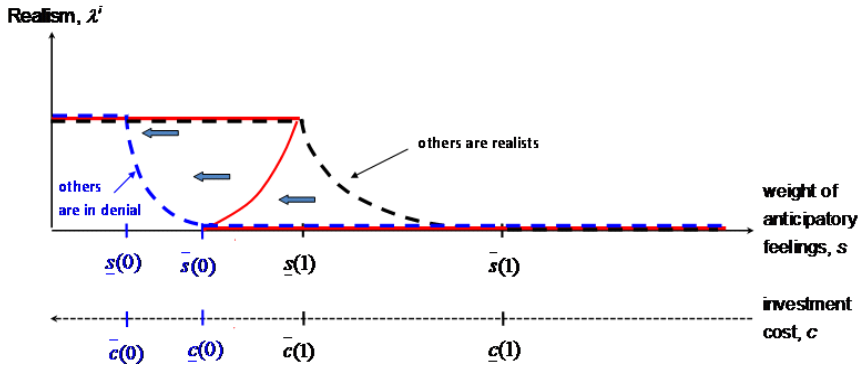

Figure 3: Groupthink ($\theta_L < 0$). The dashed lines give agent i's optimal awareness λ^i when others are realists ($\lambda^j = 1$) or deniers ($\lambda^j = 0$); arrows indicate the shift between the two. The solid lines define the social equilibria.

The intuition for what I shall term the *"Mutually Assured Delusion" (MAD)* principle is simple. If others' blindness to bad news leads them to act in a way that is better for an agent than if they were well informed ($\theta_L > 0$), it makes those news not as bad, thus reducing his own incentive to engage in denial. But if their avoidance of reality makes things worse than if they reacted appropriately to the true state of affairs ($\theta_L < 0$), future prospects become even more ominous, increasing the incentive to look the other way and take refuge in wishful thinking. In the first case, individual's ways of thinking are strategic *substitutes,* in the latter they are strategic *complements.* It is worth emphasizing that this "psychological multiplier", less than 1 in the first case and greater in the second, arises even though agents' payoffs are separable and there is no scope for social learning.

Proposition 1 shows that the scope for contagion hinges on whether overoptimism has positive or negative spillovers. Examples of both types of interaction are provided below, using financial institutions as the main illustration.

• *Limited-stakes projects, public goods:* $\theta_L > 0$. The first scenario characterizes activities with limited downside risk, in the sense that pursuing them remains socially desirable for the organization even in the low state where the private return falls short of the cost. This corresponds for instance to a bank's employees issuing "plain vanilla" mortgages or lending to safe, brick-and mortar companies –activities that remain generally profitable even in a mild recession, though less so than in a boom. Other areas in which an individual's motivation and "can-do" optimism is always valuable to others include team sports, political mobilization and other forms of good citizenship.

• *High-stakes projects:* $\theta_L < 0$. The second scenario corresponds to ventures in which the downside is severe enough that persisting has *negative social value* for the organization. The archetype is a firm like Enron, Lehman Brothers, Citigroup or AIG, whose high-risk strategy could be either extremely profitable (state H) or dangerously misguided (state L), in which case most stakeholders are likely to bear heavy losses: layoffs, firm bankruptcy, evaporated stock values, pensions and reputations, costly lawsuits or even criminal prosecution.

In such contexts, the greater is other players' tendency to ignore danger signals about "tail risk" and forge ahead with the strategy –accumulating yet more subprime loans and CDO's on the balance sheet, increasing leverage, setting up new off-the-books partnerships– the deeper and more widespread the losses will be if the scheme was flawed, the assets "toxic", or the accounting fraudulent. Therefore, when red flags start mounting, the greater is the temptation for everyone whose future is tied to the firm's fate to also look the other way, engage in rationalization, and "not think about it".[28]

The proposition's second result shows how cognitive interdependencies (of both types) are amplified, the more closely tied an individual's welfare is to the actions of others.[29] Groupthink is thus most important for closed, cohesive groups whose members perceive that they largely share *a common fate* and have few exit options. This is in line with Janis' (1972) findings, but with a more operational notion of "cohesiveness", $1-\alpha$. Such vesting can be exogenous or arise from a prior choice to join the group, in which case wishful beliefs about its future prospects also correspond to ex-post rationalizations of a sunk decision.[30]

2.3. Social cognition

I now solve for a full social equilibrium in cognitive strategies, looking for fixed points of the mapping $\lambda^{-i} \to \lambda^i$. The main intuition stems from Proposition 1 and is illustrated by the solid lines in Figures 2 and 3. From (10)-(11), $\lambda = 1$ is an equilibrium (realism is the best response to realism) for $s \leq \underline{s}(1)$, and similarly $\lambda = 0$ is an equilibrium (denial is the best

[28] Enron's employees, whose pension portfolios had on average 58% in company stock, could have moved out at nearly any point, but most never did (Samuelson (2001)). At Bears Stearns, 30% of the stock was held until the last day by employees –with presumably good access to diversification and hedging instruments– who thus lost their capital together with their job. The pattern was similar at many other financial institutions.

[29] This intuition is reflected in (9), through the term $(1-\alpha)\lambda^{-i}\theta_L$. A lower α also increases the cost of suboptimal effort when $\theta_L > 0$ and raises it when $\theta_L < 0$, reinforcing this effect (term $c - \alpha(\delta + s)\alpha\theta_L$).

[30] Such a prior investment stage is modeled in Section 5, in the context of asset markets.

response to denial) for $s \geq \bar{s}(0)$, where

$$\underline{s}(1) = \frac{m/\delta + c - \delta\alpha\theta_L}{\theta_H}, \tag{12}$$

$$\bar{s}(0) = \frac{m/\delta + c - \delta\alpha\theta_L}{\alpha\theta_L + q\Delta\theta}. \tag{13}$$

When $\theta_L > 0$ (cognitive substitutes), $s(\lambda^{-i})$ and $\bar{s}(\lambda^{-i})$ are both decreasing in λ^{-i}, so $\underline{s}(1) < \bar{s}(1) < \bar{s}(0)$ and the two pure equilibria correspond to distinct ranges. When $\theta_L < 0$ (cognitive complements), on the other hand, both thresholds are increasing in λ^{-i}, and if that effect is strong enough one can have $\bar{s}(0) < \underline{s}(1)$, creating a range of overlap.

Proposition 2. ***(Groupthink)*** *(i) If the following condition holds,*

$$(1-q)(\theta_H - \theta_L) < (1-\alpha)(-\theta_L), \tag{14}$$

then $\bar{s}(0) < \underline{s}(1)$ and for any s in this range, both realism ($\lambda = 1$) and collective denial ($\lambda = 0$) are equilibria, with an unstable mixed-strategy equilibrium in between. Under denial agents always choose $e^j = 1$, even when it is counterproductive.
(ii) If (14) is reversed, $\underline{s}(1) < \bar{s}(0)$. The unique equilibrium is $\lambda = 1$ to the left of $(\bar{s}(1), \underline{s}(0))$, a declining function $\lambda(s)$ inside the range, and $\lambda = 0$ to the right of it.
(iii) The same results characterize the equilibrium set as a function of c, with a nonempty range of multiplicity $[\bar{c}(1), \underline{c}(0)]$ if and only if (14) holds.

Equation (14) reflects the MAD principle at work. The left-hand side is the basic incentive to think that actions are highly productive (θ_H rather than θ_L) when there are no spillovers ($\alpha = 1$) or, equivalently, fixing everyone else's behavior at $e = 1$ in both states. The right-hand side corresponds to the expected losses –relative to what the correct course of action would yield– inflicted on an agent by others' delusions, and which he can (temporarily) avoid recognizing by denying the occurrence of the bad state altogether. These endogenous losses, which *transform reality from second best* to *third best,* must be of sufficient importance relative to the first, unconditional, motive for denial.

• *Comparative statics.* The proposition also yields several testable predictions. First, there is the stark reversal in how agents respond to others' beliefs (or actions) depending

on the sign of θ_L. Second, complete comparative statics on the equilibrium set are obtained. Focusing on the more interesting case where (14) holds:

(a) The more vested in the group outcome are its members, the more likely is collective denial –a form of *escalating commitment*: as $1-\alpha$ increases, both $\bar{s}(0)$ and $\underline{s}(1)$ decrease (since $\theta_L < 0$) and therefore so do the highest and lowest equilibrium values of λ. In particular, it is easy to find (Corollary 1 in the Appendix) a range of parameters for which an isolated agent *never* self-deceives, but when interacting with others, all of them *always* do so.

(b) A more desirable high state θ_H has the same effects. A more likely one (higher q) also lowers the equilibrium threshold for $\lambda = 0$, but leaves that for $\lambda = 1$ unchanged; consequently, it expands the range where multiplicity occurs.

(c) A worse low state θ_L has two effects. First, the private cost of a wrong decision rises, making a realistic equilibrium easier to sustain as there is no harmful delusion of others to "escape from": $\underline{s}(1)$ increases. When others are in denial, however, a lower θ_L also worsens the damage they do.[31] If $1/\alpha - 1/q$ is small this effect is dominated by the previous one, so $s(0)$ increases: sufficiently bad news will force people to "snap out" of collective delusion. With closely tied fates or high priors ($1/\alpha - 1/q$ large enough), on the other hand, the "scaring" effect dominates. Thus $\bar{s}(0)$ decreases, the range of multiplicity widens, and a worsening of bad news can now cause a previously realistic group to take refuge in groupthink.

• *Implications.* The types of enterprises most prone to collective delusions are thus:

(a) Those involving new and complex technologies or products that combine a generally profitable upside with a lower-probability but potentially disastrous downside –a "black swan" event. High-powered incentives, such as performance bonuses affected by common market uncertainty, have similar effects, as do highly leveraged investments that put the firm at risk of bankruptcy.

(b) Those in which participants have only *limited exit options* and, consequently, a lot riding on the soundness or folly of other's judgements. Such dependence typically arises from irreversible or illiquid *prior investments:* specific human capital, company pension plan, professional reputation, etc. Alternatively, it could reflect the *large-scale* public good nature of the problem: state of the economy, quality of the government or other society-wide

[31] From (13), $\text{sgn}\{\partial \bar{s}(0)/\partial \theta_L\} = \text{sgn}\{1/\alpha - 1/q - \delta\theta_H/(m/\delta + c)\}$, with $1/\alpha - 1/q > 0$ by (14).

institutions which a single individual has little power to affect, global warming, etc.[32]

Finally, the model shows how a propensity to "can-do" optimism (high s) can be very beneficial at the entrepreneurial stage –starting a business, mobilizing energies around a new project ($\theta_L > 0$)– but turn into a source of danger once the organization has grown and is involved in more high-stakes ventures (e.g., a mean-preserving spread in θ, with $\theta_L < 0$).[33]

2.4. Asymmetric roles: hierarchies and corporate culture

I now relax all symmetry assumptions, as well as the state-invariance of payoffs to "inaction" ($e = 0$). I then use this more general framework to show how, in hierarchical organizations, cognitive attitudes will "trickle down" and subordinates follow their leaders into realism or denial. Let the payoff structure (1) be extended to

$$U_2^i \equiv \sum_{j=1}^{n} \left(a_\sigma^{ji}\ e^j + b_\sigma^{ji}\ (1 - e^j) \right), \quad \text{for all } i = 1, \dots n \text{ and } \sigma \in \{H, L\}\,. \tag{15}$$

Each agent j's choice of $e^j = 1$ thus creates a state-dependent value a_σ^{ji} for agent i, while $e^j = 0$ generates value b_σ^{ji}; for $i = j$, these correspond to agent i's private returns to action and inaction. All payoffs remain linearly separable for the same expositional reason as before, but complementarities or substitutabilities are easily incorporated (see Section 2.5). Agents may also differ in their preference and cognitive parameters c^i, m^i, δ^i, their proclivity to anticipatory feelings s^i or even their priors q^i. The generalization of (3) is then

$$a_L^{ii} - b_L^{ii} < \frac{c^i}{s^i + \delta^i} < q^i \left(a_H^{ii} - b_H^{ii} \right) + (1 - q^i) \left(a_L^{ii} - b_L^{ii} \right), \tag{16}$$

while that of $\theta_H > \theta_L$ (H is the better state under full information), is

$$\sum_{j=1}^{n} a_H^{ji} > \sum_{j=1}^{n} b_L^{ji}. \tag{17}$$

[32] This point is pursued in Bénabou (2008), where I study the dynamics of country-level ideologies concerning the relative efficacy of markets and governments.

[33] Similarly, through most of human history collective activities (hunting, foraging, fighting, cultivation) were typically characterized by $\theta_L > 0$, making group morale valuable and susceptibility to optimism (a high s or low m) an evolutionary advantageous trait. (For a related account, see von Hippel and Trivers (2011)). Modern technology and finance now involve many high-stakes activities ($\theta_L << 0 << \theta_H$), for which those same traits can be a source of trouble. With leverage, for instance, payoffs become $\theta'_H \equiv \theta_H + B(\theta_H - R)$ and $\theta'_L \equiv \theta_L + B(\theta_L - R)$, where B is borrowing and $R \in (\theta_L, \theta_H)$ the gross interest rate.

Following the same steps as in the symmetric case and denoting Λ^{-i} the vector of other agents' strategies, it is easily seen that agent i's best response λ^i is similar to that in Lemma 1, but with the cutoffs for realism and denial now given by

$$\underline{s}^i(\Lambda^{-i}) \equiv \frac{m^i/\delta^i + c^i - \delta^i\left(a_L^{ii} - b_L^{ii}\right)}{\Sigma_{j=1}^n\left(a_H^{ji} - a_L^{ji}\right) + \Sigma_{j\neq i}\,\lambda^j\left(a_L^{ji} - b_L^{ji}\right) + a_L^{ii} - b_L^{ii}}, \tag{18}$$

$$\overline{s}^i(\Lambda^{-i}) \equiv \frac{m^i/\delta^i + c^i - \delta^i\left(a_L^{ii} - b_L^{ii}\right)}{q\left[\Sigma_{j=1}^n\left(a_H^{ji} - a_L^{ji}\right) + \Sigma_{j\neq i}\,\lambda^j\left(a_L^{ji} - b_L^{ji}\right)\right] + a_L^{ii} - b_L^{ii}}. \tag{19}$$

Thus λ^i is (weakly) increasing in λ^j, representing cognitive *complementarity,* whenever $a_L^{ji} - b_L^{ji} < 0$, meaning that j's delusions (leading to $e^j = 1$ when $\sigma = L$) are harmful to i; conversely, $a_L^{ji} - b_L^{ji} > 0$ leads to *substitutability.* This is a bilateral version of the MAD principle. Similarly, agent i is more likely to engage in denial when surrounded by deniers ($\lambda^j \equiv 0$) than by realists ($\lambda^j \equiv 1$) if and only if $\sum_{j=1}^n (a_L^{ji} - b_L^{ji}) < 0$, meaning that others' mistakes are harmful *on average,* and generalizing $\theta_L < 0$. Multiple equilibria occur when this (expected) loss is sufficiently large relative to the "unconditional" incentive to deny:

$$(1-q)\sum_{j=1}^{n}\left(a_H^{ji} - a_L^{ji}\right) < \sum_{j\neq i}\left(b_L^{ji} - a_L^{ji}\right), \tag{20}$$

which clearly generalizes (14).

Proposition 3. ***(Organizational cultures)*** *Let (16)-(20) hold for all $i = 1, \ldots n$. There exists a non-empty range $[\overline{s}^i(0), \underline{s}^i(1)]$ (respectively, $[\overline{c}^i(1), \underline{c}^i(0)]$) for each i, such that if $(s^1, \ldots s^n) \in \Pi_{i=1}^n [\overline{s}^i(0), \underline{s}^i(1)]$ (respectively, if $(c^1, \ldots c^n) \in \Pi_{i=1}^n [\overline{c}^i(1), \underline{c}^i(0)]$) both collective realism ($\lambda^i \equiv 1$) and collective denial ($\lambda^i \equiv 0$) are equilibria.*[34]

• *Directions of cognitive influence.* Going beyond multiplicity, interesting results emerge for organizations in which members play asymmetric roles. Thus, (18)-(19) embody the intuition that an agent's way of thinking is most sensitive to how the people whose decisions have the greatest impact on his welfare (in state L) deal with unwelcome news:

$$\frac{\partial \underline{s}^i}{\partial \lambda^j} >> \left|\frac{\partial \underline{s}^j}{\partial \lambda^i}\right| \text{ and } \frac{\partial \overline{s}^i}{\partial \lambda^j} >> \left|\frac{\partial \overline{s}^j}{\partial \lambda^i}\right| \quad \text{iff} \quad \frac{b_L^{ji} - a_L^{ji}}{\left|b_L^{ij} - a_L^{ij}\right|} >> \max\left\{\left(\frac{\underline{s}^j}{\underline{s}^i}\right)^2, \left(\frac{\overline{s}^j}{\overline{s}^i}\right)^2\right\}. \tag{21}$$

[34] As usual, there is also an odd number of mixed-strategy equilibria in-between. I do not focus on these, as they are complicated to characterize (especially with asymmetric agents) and do not add any insight.

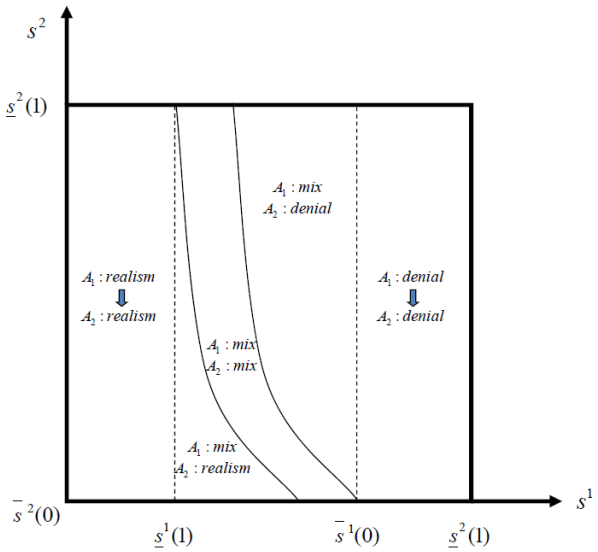

Figure 4: "Trickle down" of realism and denial in a hierarchy. The equilibrium strategies of manager (A_1) and worker(s) (A_2) are indicated in each region, with the arrows illustrating complete top-down determination.

This condition is ensured in particular when $\left|a_L^{ij} - b_L^{ij}\right| << |a_L^{ii} - b_L^{ii}|$ and

$$b_L^{ji} - a_H^{ji} >> \max\left\{\sum_{k\neq i,j}\left|a_L^{ki} - b_L^{ki}\right|,\ \left|a_L^{ii} - b_L^{ii}\right|, \sum_{j=1}^{n}\left|a_H^{ji} - a_L^{ji}\right|\right\}. \tag{22}$$

Consider, for instance, the simplest form of hierarchy: two agents, 1 and 2, such as a manager and worker. If $a_L^{12} - b_L^{12}$ is sufficiently negative while $|a_L^{21} - b_L^{21}|$ is relatively small, agent 2 suffers a lot when agent 1 loses touch with reality, while the converse is not true. Workers thus risk losing their job if management makes overoptimistic investment decisions, whereas the latter has little to lose if workers put in more effort than realistically warranted. When the asymmetry is sufficiently pronounced it leads to a testable pattern of predominantly *top-down cognitive influences,* illustrated in Figure 4.

Proposition 4. ***(Cognitive trickle-down)*** *There exists a nonempty range of parameters such that* $[\underline{s}^1(1), \bar{s}^1(0)] \subset [\bar{s}^2(0), \underline{s}^2(1)] \equiv S$ *and, for all* $(s^1, s^2) \in S \times S$*, the equilibrium is unique and such that:*

(i) The qualitative nature of the manager's cognitive strategy –complete realism, complete denial, or mixing– depends only on her own s^1*, not on the worker's* s^2*.*

(ii) If the manager behaves as a systematic denier (respectively, realist), so does the worker:

where $\lambda^1 = 1$ *it must be that* $\lambda^2 = 1$, *and similarly* $\lambda^1 = 0$ *implies* $\lambda^2 = 0$.
(iii) Only when both agents are in partial denial (between the two curves in Figure 4) does the worker's degree of realism also influence that of the manager.

Let agent 2 now be replicated into $n-1$ identical workers, each with influence $[a_\sigma^{j1} e^j + b_\sigma^{j1}(1-e^j)]/(n-1)$ over the manager, but subject to the same influence from him as before, $a_\sigma^{1j} e^1 + b_\sigma^{1j}(1-e^1)$. Figure 4 then remains operative, showing how *the leader's attitude toward reality* tends to *spread to all his subordinates,* while being influenced by theirs only in a limited way, and over a limited range.

This result has clear applications to corporate and bureaucratic culture, explaining how people will contagiously invest *excessive faith in a leader's "vision".*[35] Likewise in the political sphere, a dictator need not exert constant censorship or constraint to implement his policies, as crazy as they may be: he can rely on people's mutually reinforcing tendencies to rationalize as "not so bad" the regime they (endogenously) have to live with.

The above is of course an oversimplified representation of an organization; yet the same principles will carry over to more complex hierarchies with multiple tiers (by "chaining" condition (21) across levels i, j, k, etc.), strategic interactions, control rights, transfer payments, etc. Such extensions lie outside the scope of this paper and are left to future work.

2.5. Robustness

While the benchmark model developed in this section involves a number specific assumptions, the insights it delivers are very general. This subsection (which can be skipped) explains how the main results extend to a series of increasingly different settings.

[35] In Rotemberg and Saloner (1993), a manager's "vision" (prior beliefs or preferences that favor some activities over others) serves as a commitment device to reduce workers' concerns about ex-post expropriation of their innovations. In Prendergast (1993), managers' use of subjective performance evaluations to assess subordinates' effort at seeking new information leads the latter to distort their reports in the direction of the manager's (expected) signal. Both mechanisms thus lead workers to conform their behavior to managers' prior beliefs. Unlike here, however, in neither case do they actually espouse those beliefs, nor would the manager ever want them to report anything but the truth. In Hermalin (1998), a leader with private information about the return to team effort works extra-hard to convince his coworkers to do so; the resulting separating equilibrium shifts up the whole profile of efforts (ameliorating the free-rider problem) but involves no mistaken belief by anyone. Manager and workers also share beliefs in Van den Steen (2005) but, rather than via learning, this arises from agents with diverse priors sorting themselves through the labor market. Managers with a strong "vision" thus tend to attract employees with similar priors, as this helps alleviate incentive and coordination problems within the firm.

• *Strategic interactions.* The focus has so far been on environments in which an agent's welfare depends on others' actions, but his return to acting does not. Quite intuitively, strategic complementarities in payoffs will reinforce the tendency for contagion, whereas substitutabilities will work against it.[36] To see this, let agent i' expected payoff in state $\sigma = H, L$ now be $\Pi^i_\sigma(e^i, \mathbf{e}^{-i})$, where $\mathbf{e}^{-i}$ denotes the vector of others' actions; his incentive to act is then $\pi^i_\sigma(\mathbf{e}^{-i}) \equiv \Pi^i_\sigma(1, \mathbf{e}^{-i}) - \Pi^i_\sigma(0, \mathbf{e}^{-i})$. In state L, the differential in i's anticipatory value of denial that results from others' "blind" persistence, previously given by $-s(1-\alpha)\theta_L$, is now $-s[\Pi^i_L(1, \mathbf{0}) - \Pi^i_L(1, \mathbf{1})]$, which embodies the same MAD intuition. The new ingredient is that others' persistence now also changes the return to investing in state L (previously a fixed $\alpha\theta_L$), by $\pi^i_L(\mathbf{1}) - \pi^i_L(\mathbf{0})$, with sign governed by $\Sigma_{j \neq i}\ \partial^2 \Pi^i_L / \partial e^i \partial e^j$. When actions are complements, delusion is thus less costly if others are also in denial, whereas with substitutes (as in the asset market of Section 5) it is more costly.

• *Signal structure.* Instead of "tuning out" bad news, selective awareness can take the form of spending resources to retain good ones –through rehearsal, preserving evidence, etc. This case, in which attention or recall is naturally imperfect but can be raised at some cost, is equivalent to setting $m < 0$, with all key results unchanged. The use of binary signals and actions is also inessential: with a richer state space, self-deception takes the form of a partitional coarsening of signals, as is standard in models of communication.

• *Sophistication.* While the model is an equilibrium one, strategic sophistication and common knowledge of rationality are inessential to the main results. For denial to be contagious, for instance, an agent does not need to know *why* others around him are escalating a risky corporate strategy, or accumulating dubious assets (Section 5) in spite of mounting red flags. It suffices that he see that they do ($e^j = 1$ when $\sigma = L$) and simply understand that this worsens his prospects: greater leverage implies a higher probability of firm bankruptcy if profits fall short, greater market buildup a deeper crash if fundamentals are weak, etc. Formally, the key property is that the slope of an agent's cognitive best-response (λ^i) to others' material actions (e^{-j}) in state L hinges on whether he is made worse or better or worse

[36] Sources of complementarity may include technological gains from coordination or a desire for social conformity –whether intrinsic or resulting from sanctions imposed on norm violators. At the same time, without anticipatory feelings, preferences for late resolution of uncertainty or some other non-standard role of beliefs, no amount of complementarity can generate results similar to the model's: agents with standard preferences, including "social" ones, always have (weakly) positive demand for knowledge and thus never engage in reality denial or information avoidance.

by their mistakes (e.g., the sign of θ_L). A bounded-rationality version of the model, in which agents simply best-respond to past aggregate investment, is thus shown in the Appendix to yield results very similar to those of the fully rational case.[37]

• *Preferences and cognition.* Most importantly, the model's findings on cognitive influences (complements and substitutes, horizontal and vertical) and their determinants are *entirely independent* of the assumptions of anticipatory utility and malleable memory used to represent individual belief distortion. As shown below, replacing this specification with Kreps-Porteus (1978) preferences leads to closely related (but also complementary) results. More generally, the MAD idea provides a portable template for belief contagion that could be applied to any individual-decision model generating informational preferences.

3. Contagious ignorance: the role of risk

In this section I derive versions of the MAD principle and groupthink results that are based on intertemporal risk attitudes rather than anticipatory utility, and where willful blindness takes the form of *ex-ante* information avoidance (not wanting to know) rather than *ex-post* belief distortion (reality denial). There are three reasons for doing so. First, as seen earlier, both types of behaviors are observed in experiments and real-world situations. Second, the role of risk in cognitive distortions is of intrinsic interest. Finally, this will make clear that the paper's results are not tied to any particular assumption about the individual motive for non-standard updating, nor the form that the latter takes. They concern instead the *social transmission of beliefs,* which a simple and general insight relates to the structure of interactions among agents. In the present case, it implies that willful ignorance will be contagious (complementarity) when its collateral effect is to *magnify the risks* borne by others, and self-dampening (substitutability) when it *attenuates* those risks.

• *Technology.* I maintain the general interaction structure of Section 2.4, which will bring to light most clearly the roles played by different types of risks.[38] For simplicity, all payoffs are now received in the last period ($t = 2$), with[39]

[37] The same would be true with other standard specifications of adaptive learning, such as fictitious play (e.g., Fudenberg and Levine (1998)) or replicator dynamics.

[38] In the restricted symmetric model of Section 2.1, by contrast, parameters such as θ_L or α affect both the variance and mean of payoffs. Thus, while results qualitatively similar to those of Proposition 6 can be obtained, they are not easily interpretable and the conditions required are much more constraining.

[39] Any costs incurred in period 1 are thus "folded into" the final payoffs, with appropriate discounting:

$$a_H^{ii} - b_H^{ii} > 0 > a_L^{ii} - b_L^{ii} \equiv -f_L^i, \tag{23}$$

$$qa_H^{ii} + (1-q)a_L^{ii} > qb_H^{ii} + (1-q)b_L^{ii}, \tag{24}$$

$$d_L^i \equiv \sum_{j \neq i} \left(b_L^{ji} - a_L^{ji}\right) \gtrless 0, \tag{25}$$

$$A_H^i \equiv \sum_{i=1}^{n} a_H^{ji} \gtrless \sum_{i=1}^{n} b_L^{ji} \equiv B_L^i. \tag{26}$$

The first equation specifies that the privately optimal action for agent i is $e^i = 1$ in state H and $e^i = 0$ in state L. The second one implies that when uninformed, a risk-neutral agent will choose $e^i = 1$; if the state turns out to be L, he then incurs a loss of $f_L^i > 0$ (f stands for "fault"). The third equation defines the total impact on agent i that results when everyone else chooses $e^j = 1$ in state L, which they will do if uninformed. The most natural case is that where $d_L^i \geq 0$ (so d stands for collateral "damage"), but I also allow $d_L^i < 0$. The last equation compares which of state H or L is better for agent i when everyone is informed; the most plausible case is $A_H^i > B_L^i$, but this is not required for any of the results.

• *Preferences.* I simply replace the combination of anticipatory preferences and malleable memory used so far with Kreps-Porteus (1978) preferences. Thus, at date 1 agents evaluate final lotteries according to an expected utility function $U_1 = E_1[u(x)]$, and at date 0 they evaluate lotteries over date-1 utilities U_1 according to an expected utility function $E_0\left[v(U_1)\right]$. Expectations are now standard rational forecasts (there is no forgetting) and agents' only informational choice is *whether or not to learn* the signal $\sigma = H, L$ at $t = 0$. Both options are taken to be costless, but it would be trivial to allow for positive costs of becoming informed or remaining uninformed. For comparability with the previous results I take agents to be risk-neutral at date 1, $u(x) \equiv x$. The function $v(x)$, on the other hand, is strictly concave, generating a *ceteris paribus* preference for the *late resolution of uncertainty.* To avoid corner solutions I take $v(x)$ to be defined over all of $\mathbb{R}$, and for some results will also require (without much loss of generality) that that there exist $\gamma > 1$ and $\gamma' > 1$ such that[40]

$$\lim_{x \to +\infty} [v(x)/x^{1/\gamma}] \quad \textit{and} \quad \lim_{x \to -\infty} [-v(x)/\left(-x\right)^{\gamma'}] \quad \textit{are well-defined and positive.} \tag{27}$$

thus a_H^{ii} corresponds here to $a_H^{ii} - c^i/\delta^i$ in Section 2.4.

[40] For instance, $v(x) = 1 - \gamma + \gamma\left(x+1\right)^{1/\gamma}$ for $x \geq 0$, $v(x) = 2 - \left(1 - x/\gamma\right)^{\gamma}$ for $x \leq 0$. More generally, any strictly increasing and concave function $v(x)$ defined on $\mathbb{R}_+$, with $0 < v'(0) < +\infty$ and $v(+\infty) = +\infty$ can be extended by symmetry around the perpendicular to its tangent at $(0, v(0))$: for all $x \leq 0$, $v(x) \equiv v(0) - v'(0)v^{-1}(v(0) - v'(0)x)$. The assumptions $\text{supp}(v) = \mathbb{R}$ and (27) could also be substantially weakened.

At $t = 0$, when deciding whether or not to learn the state of the world, agents face a tradeoff between their preference for late resolution and the decision value of information. The novel feature of the problem considered here is that each one's prospects also depend on how *others* act, and therefore on who else chooses to be informed or remain ignorant.

• *The MAD principle for risks.* Consider an agent i and let $d \in \mathbb{R}$ parametrize the losses he will incur due to the mistakes of those who choose $e^j = 1$ in state L. Thus $d = \Sigma_{j \in J} \left(b_L^{ji} - a_L^{ji}\right) \gtrless 0$, where J denotes the uninformed subset. Agent i' final payoffs are given by the lottery $\mathcal{I}(d)$ if he finds out the state at $t = 0$ and by $\mathcal{N}(d)$ if he does not, where:[41]

$$\mathcal{I}(d) \equiv \begin{cases} q: & A_H^i \\ 1-q: & B_L^i - d \end{cases}, \quad \mathcal{N}(d) \equiv \begin{cases} q: & A_H^i \\ 1-q: & B_L^i - f_L^i - d \end{cases}. \tag{28}$$

He therefore prefers to remain ignorant if

$$\varphi^i(d) \equiv v\left(qA_H^i + (1-q)\left(B_L^i - f_L^i - d\right)\right) - qv(A_H^i) - (1-q)\, v(B_L^i - d) > 0. \tag{29}$$

Consider first the case in which everyone else is informed or, equivalently, agent i is insulated from their mistakes.[42] Thus $d = 0$, and he prefers to know the state if

$$\varphi^i(0) = v\left(qA_H^i + (1-q)\left(B_L^i - f_L^i\right)\right) - qv(A_H^i) - (1-q)\, v(B_L^i) < 0. \tag{30}$$

Since v is strictly increasing, this holds when faulty decisions are costly enough,

$$f_L^i > \underline{f}^i, \tag{31}$$

where $\underline{f}^i > 0$ is defined by equality in (30).

Consider now the role of d: as it rises, (28) makes clear how others' ignorance renders agent i's future more risky, increasing the variance in *both* feasible prospects $\mathcal{I}(d)$ and $\mathcal{N}(d)$. This extra risk, which he cannot avoid, makes finding out whether the state is H or L more frightening and thus reduces his willingness to know. The following results, illustrated in Figure 5, characterize more generally each agent's attitude towards information.

[41] For simplicity, agents here have a common prior, $q^i = q$. This can easily be relaxed, as in the previous section, and so can their having the same utility function v.

[42] This corresponds in particular to a single agent facing payoffs given by (28) in which $d = 0$ and A_H^i, B_L^i represent exogenous state-contingent prizes.

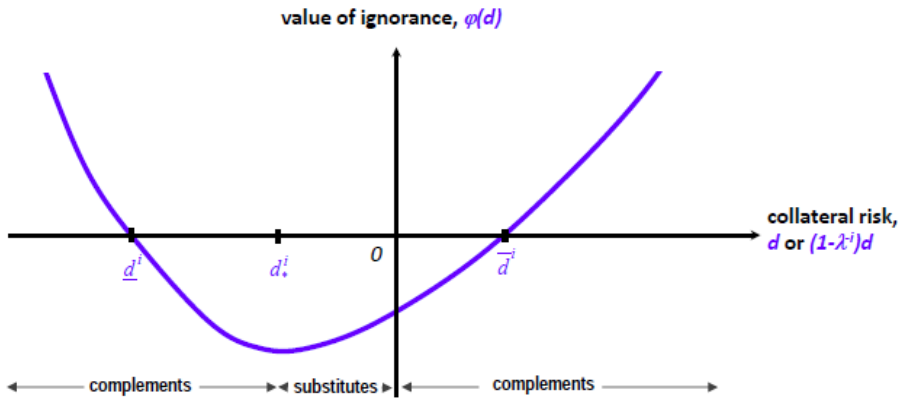

Figure 5: collateral risk and informational decisions

Lemma 2. *The function $\varphi^i(d)$ is strictly quasiconvex, reaching a negative minimum at*

$$d_*^i \equiv -\left(A_H^i - B_L^i\right) + \left(\frac{1-q}{q}\right) f_L^i, \tag{32}$$

independent of $v(\cdot)$. Furthermore, if $v(\cdot)$ satisfies (27) then $\varphi^i(d) \to +\infty$ as $|d| \to +\infty$, so there exists finite thresholds $\underline{d}^i < d_*^i < \bar{d}^i$ such that $\varphi^i(d) > 0$ if and only if $d \notin [\underline{d}^i, \bar{d}^i]$.

The intuition is clearest when d is positive and relatively large, meaning that others' mistakes impose nontrivial collateral damages in state L; this is also the most empirically relevant case. What matters is payoff *risk,* however, so information aversion also occurs when others' ignorance has a sufficiently positive payoff –that is, when d is negative enough.[43] The size of the collateral stakes $|d|$, or more precisely its contribution to $|d - d_*^i|$, plays here the same role for agents who dislike *variance* in their date-1 utility U_1^i as d itself (or $-(1-\alpha)\theta_L$ in the symmetric case) played earlier for agents disliking a low *level* of U_1^i. The term d_*^i corrects in particular for the fact that it is not just the sum of risks that matters, but also their correlation: remaining uninformed leads to a costly mistake (f_L^i) when L occurs, which is also when the agent incurs d from others' ignorance.[44]

These results lead a to full characterization of agents' cognitive best responses.

[43] Note also that $\left(\varphi^i\right)'(d) > 0$ on $\mathbb{R}_+$ as long as $d_*^i < 0$, or equivalently $qA_H^i + (1-q)\left(B_L^i - f_L^i\right) > B_L^i$. This condition is most plausible, as it means that a single risk-neutral agent at date 1 prefers the lottery $\mathcal{N}(0)$ to the degenerate one in which the state is L with probability 1. In the benchmark model of Section 2.1, for instance, $A_H^i = \theta_H - c^i/\delta^i$, $B_L^i = 0$ and $f_L^i = c^i/\delta^i - \alpha\theta_L$, so $d_*^i < 0$ is always implied by (24).

[44] This increases the value of information for $d > 0$ and lowers it for $d < 0$, thus raising the threshold d_*^i beyond which higher d's makes the agent less willing to become informed ($\varphi' > 0$). For $f_L^i = 0$, $|d - d_*^i| = |A_H^i - (B_L^i - d)|$ is just the spread in payoffs common to $\mathcal{I}(d)$ and $\mathcal{N}(d)$. Note also how the opposite roles of avoidable and unavoidable risks are reflected in φ^i, which is concave in f_L^i and quasiconvex in d.

Proposition 5 (MAD principle for risks). *(i) Given any two subsets of agents J and J' not containing i, denote $d = \Sigma_{j\in J}\ (b_L^{ji} - a_L^{ji})$ and $d' = \Sigma_{j\in J'}\ (b_L^{ji} - a_L^{ji})$. Agent i's incentive to avoid information is higher when the set of uninformed agents is J' rather than J if and only if $(d' - d)\,(d - d_*^i) > 0$.*

(ii) Let each agent be equally affected by the mistakes of all others: $b_L^{ji} - a_L^{ji} = d$ for all i, j with $j \neq i$. The informational choices of all agents are strategic complements if d lies outside the interval $[\min\{d_*^i, 0\}, \max\{d_*^i, 0\}]$, and strategic substitutes if it lies within.

The first part of the proposition demonstrates the role of collateral risk most generally. First, if $J \subset J'$, more agents remaining ignorant make i more averse to information when they add to the total risk he bears, in the sense of moving d further away from d_*^i. Second, taking J and J' disjoint (for example, i's hierarchical superiors and subordinates, respectively) shows that an agent's wanting or not wanting to know is most sensitive to how the people whose ignorance imposes the greatest risk on him deal with uncertainty. This naturally leads, as in Section 2.4, to a trickle-down of attitudes towards information –from management to workers, political leader to followers, etc.

The second part of the proposition is illustrated in Figure 5 by a simple rescaling of d. In this "horizontal" case the value of ignorance is $\varphi^i((1 - \lambda^{-i})d)$, where $1 - \lambda^{-i}$ is the fraction of others who choose to remain uninformed and d is now the "normalized" damage.

• *Groupthink as contagious ignorance.* When the total uncertainty he faces due to the ignorance of others ($d = d_L^i$ defined in (25)) is large enough, an agent who would otherwise have positive demand for information ($f_L^i > \underline{f}^i$) will prefer to also avoid learning the state of the world. Thus $\varphi^i(0) < 0 < \varphi^i(d_L^i)$, meaning that knowledge is a best reply to knowledge and ignorance a best reply to ignorance, in a manner that echoes Propositions 2 and 3. As a consequence, *risk also spreads* and becomes systemic throughout the organization.

Proposition 6 (endogenous systemic risk). *Let (23), (24) and (31) hold for all i, and $v(\cdot)$ satisfy (27). There exists a non-empty set $D^i \equiv (-\infty, \underline{d}^i) \bigcup (\bar{d}^i, +\infty)$ for each i, with $\underline{d}^i < 0 < \bar{d}^i$, such that if $(d_L^1, \ldots d_L^n) \in \prod_{i=1}^{n} D^i$, for all i, both collective realism (every agent becoming informed at date 0) and collective willful ignorance (every agent choosing to remain uniformed) are equilibria. In the latter, each agent i's willingness to pay to avoid information is positive and increasing in $|d_L^i|$ on each side of D^i.*

• *The role of risk preferences.* Given a structure of interactions, intuition suggests that for multiple regimes to arise, agents' preference for late resolution should be neither too large nor too small. Indeed, if (29) (respectively, (30)) holds for some function v, it also holds for any w that is increasing and more (respectively, less) concave.[45]

Proposition 7. *Let $\{v_\gamma(x), \gamma \geq 1\}$ be a family of concave functions on $\mathbb{R}$ such that $v_{\gamma'}$ is strictly more concave than v_γ whenever $\gamma' > \gamma$. Given a payoff structure $(a_\sigma^{ij}, b_\sigma^{ij})_{\sigma=H,L}^{i,j=1,...n}$ satisfying (23)-(26), there exists a range $[\underline{\gamma}, \bar{\gamma}]$ such that the informed and uniformed organizational equilibria coexist if and only if $\gamma \in [\underline{\gamma}, \bar{\gamma}]$.*[46]

The bounds $\underline{\gamma}$ and $\bar{\gamma}$ can be derived explicitly in the case of quadratic utility: $v(x) = x - \gamma x^2/2$ for $x \in (-\infty, 1/\gamma)$. Conditions (29) and (30) then become

$$\frac{2f_L^i}{\gamma} < q\left(A_H^i - B_L^i + d_L^i + f_L^i\right)^2 - f_L^i\left(f_L^i - 2B_L^i + 2d_L^i\right), \tag{33}$$

$$\frac{2f_L^i}{\gamma} > q\left(A_H^i - B_L^i + f_L^i\right)^2 - f_L^i\left(f_L^i - 2B_L^i\right), \tag{34}$$

which respectively define $\underline{\gamma}$ and $\bar{\gamma}$. Proposition 11, given in the Appendix, shows that $\underline{\gamma} < \bar{\gamma}$ and a range of equilibrium multiplicity exists, provided $|d_L^i|$ is large enough.

• *Modeling choices.* Compared to anticipatory utility and imperfect recall, Kreps-Porteus preferences have the advantage of well-established axiomatic foundations. On the other hand, the results they lead to are much less tractable analytically. The thresholds determining equilibrium do not generally admit closed-form solutions, whereas in Propositions 1-3 they were obtained explicitly, with readily interpretable comparative statics. It may also be quite difficult for an agent to avoid informative signals, especially in a social context, so the relevant question is more often how to deal with the information one does have.

From here on I therefore *revert to the benchmark specification* of Section 2.1, simply noting that for each application one could again derive parallel results based on risk attitudes.

[45] By definition, w is more concave than v if $w = \omega \circ v$, for some increasing and concave function ω.

[46] For arbitrary v_γ's and parameter configurations, the interval could be empty ($\underline{\gamma} > \bar{\gamma}$). By Lemma 2, sufficient conditions for $\underline{\gamma} < \bar{\gamma} \leq +\infty$ are that: (i) v_γ satisfy the asymptotic conditions (27) for at least one value of γ; (ii) for this v_γ, the threshold $\underline{f}^i$ is less than f_L^i, for all i; (iii) $|d_L^i - d_*^i|$ is large enough, for all i.

4. Welfare, Cassandra's curse and free speech protections

Are members of a group in collective denial worse or better off than if they faced the truth –as an alternative equilibrium or by means of some collective commitment mechanism? I adopt here the ex-ante, behind-the-veil perspective of organizational designers who could choose the structure of payoffs (activities, incentives, employees' types) and information (hard or soft signals, treatment of dissenters) to maximize total surplus. Computing welfare as of $t = 0$ is also consistent with a revealed-preferences approach: from agents' willingness-to-pay to ensure collective realism or denial, inferences can be made about their deep preferences parameters, such as s.[47]

Consider first state $\sigma = L$. When agents are realists (setting $\lambda^j = 1$ in (7)), equilibrium welfare is $U^*_{L,R} = 0$. When they are deniers (setting $\lambda^j = 0$ in (8)), it is given by:

$$U^*_{L,D}/\delta = -m - c + \delta\theta_L + sq\theta_H + s\,(1-q)\,\theta_L. \tag{35}$$

As illustrated in Figure 6, whether collective denial of bad news is harmful or beneficial thus depends on whether s lies below or above the threshold

$$s^* \equiv \frac{m/\delta + c - \delta\theta_L}{q\theta_H + (1-q)\,\theta_L}.^{48} \tag{36}$$

Proposition 8. *Welfare following bad news (state L):*
(1) If $\theta_L < 0$, then $s^* > \max\{\bar{s}(0), \underline{s}(1)\}$. Whenever realism ($\lambda = 1$) is an equilibrium, it is superior to denial ($\lambda = 0$). Moreover, there exists a range in which realism is not an equilibrium but, if it can be achieved through collective commitment, yields higher welfare.
(2) If $\theta_L > 0$, then $s^* < \bar{s}(0)$. The equilibrium involves excessive realism for $s \in (s^*, \bar{s}(0))$ and excessive denial for $s \in (\underline{s}(1), s^*)$, when this interval is nonempty.

Given how damaging collective delusion is in state L with $\theta_L < 0$, it makes sense that when realism can also be sustained as an equilibrium it dominates, and that when it cannot the group may try to commit to it. Conversely, with $\theta_L > 0$, *boosting morale* in state L ameliorates the *free-rider problem,* so the group would want to commit to ignoring adverse

[47] One may nonetheless ask what would change if welfare was evaluated based on U^i_1 rather than U^i_0 (though it would then not be measurable through organizational-design decisions). This turns out to make no difference, apart from a trivial parameter renormalization: see footnote 50.

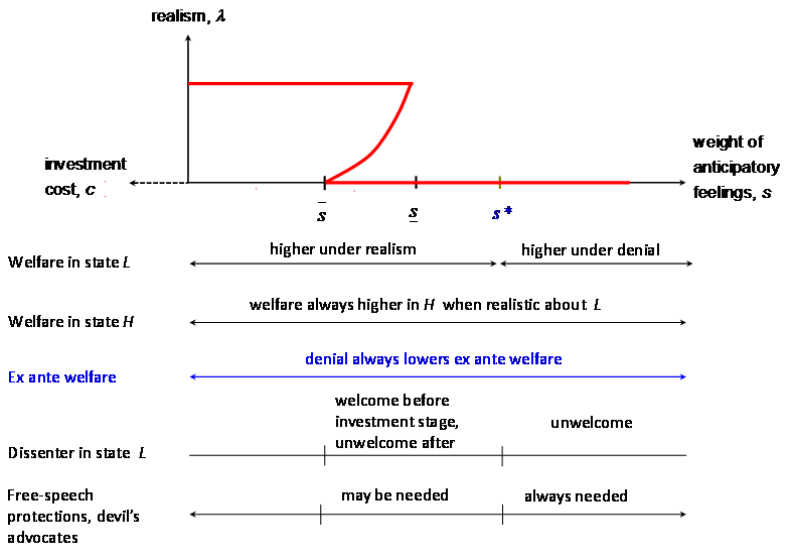

Figure 6: Welfare and dissenting speech (groupthink case)

signals when $s \geq s^*$ but the only equilibrium involves realism.[49]

Consider now welfare in state H. Given (3), everyone chooses $e^i = 1$ in both equilibria. Under denial, however, agents *can never be sure* of whether the state is truly H, or it was really L and they censored the bad news. As a result of this "spoiling" effect, welfare is only

$$U^*_{H,D}/\delta = -c + \delta\theta_H + s\left[q\theta_H + (1-q)\,\theta_L\right] < -c + (\delta + s)\,\theta_H = U^*_{H,R}/\delta. \tag{37}$$

Averaging over the two states, finally, the mean belief about θ remains fixed (by Bayes' rule), so the net welfare impact of denial, $\Delta W_0 \equiv q\left(U^*_{H,D} - U^*_{H,R}\right) + (1-q)\left(U^*_{L,D} - U^*_{L,R}\right)$, is just

$$\Delta W_0 \equiv (1-q)\delta\left[(\delta + s)\,\theta_L - c - m/\delta\right], \tag{38}$$

realized in state L. In assessing the overall value of social beliefs one can thus focus on *material* outcomes and ignore anticipatory feelings, which are much more difficult to measure but wash out across states of nature.[50]

[49] If θ_L is high enough that $\delta\theta_L > c + m/\delta$, then $s^* < 0$: overoptimism in state L is socially beneficial even absent anticipatory emotions ($s = 0$). A good example is team morale in sports.

[50] This is also true when evaluating (unconditional) utilities from the point of view of date 1. The welfare differential across denial and realistic group outcomes is then $\Delta W_1 = (1-q)\left[(\delta + s)\theta_L - c\right]$, which just amounts to renormalizing c to $c + m/\delta$ in $\Delta W_0/\delta$. Furthermore, m can be taken (if desired) as arbitrarily small or even zero; see footnote 23.

Proposition 9. *(1) Welfare following good news (state H) is always higher, the more realistic agents are when faced with bad news (the higher is λ).*
(2) If $\theta_L \leq 0$, denial always lowers ex-ante welfare. If $\theta_L > 0$, it improves it if and only if $(\delta + s)\,\theta_L > c + m/\delta$.

These results, also illustrated in Figure 6, lead to a clear distinction between two types of collective beliefs and the settings that give rise to them. They are also testable, since ΔW_0 measures agents' willingness to pay (positive or negative) for organizational designs or commitment devices that ensure collective realism.

• *Beneficial group morale.* When $\theta_L > 0$, $e = 1$ is socially optimal even in state L, but since $\alpha(s+\delta)\theta_L < c$ it is not privately optimal. If agents can all manage to ignore bad news at relatively low cost, either as an equilibrium or through commitment, they will be better off not only ex-post but also ex-ante: $\Delta W_0 > 0$. This is in line with a number of recent results showing the functional benefits of overoptimism (achieved through information manipulation or appropriate selection of agents by a principal) in settings where agents with the correct beliefs would underprovide effort.[51]

• *Harmful groupthink.* The novel case is the one in which contagious delusions can arise, $\theta_L < 0$, and it also leads to a more striking conclusion: not only can such reality avoidance greatly damage welfare in state L, but even when it improves it those gains are always dominated by the losses induced in state H : $\Delta W_0 < 0$.[52] This normative result also has positive implications for how organizations and polities deal with dissenters, revealing an important form of *time inconsistency* between ex ante and ex post attitudes.

• *The curse of Cassandra.* Let $\theta_L < 0$ and consider a denial equilibrium, as in Figure 6. Suppose now that, in state L, an individual or subgroup with a lower s or different payoffs attempts to bring the bad news back to everyone's attention. If this occurs after agents have sunk in their investments it simply amounts to deflating expectations in (2),

[51] In a team or firm context see, e.g., Bénabou and Tirole (2003), Fang and Moscarini (2005), Van den Steen (2005) and Gervais and Goldtstein (2007). In a self-control context see Bénabou and Tirole (2002) and in an intergenerational context see Dessi (2008).

[52] The "shadow of doubt" cast over the good state by the censoring of the bad state could also distort some decisions in state H, given more than two action choices. If, on the other hand, agents are less than fully aware of their own tendency to self deception, the losses in state H are attenuated and ex-ante gains become possible. Thus, with $\chi < 1$ in (6), q is simply replaced by $q/\left[q + \chi(1-q)\right]$ in (35) and (37), and ΔW_0 consequently augmented by $s\delta(1-\chi)q(1-q)/\left[q + \chi(1-q)\right]$.

so they will refuse to listen, or may even try to "kill the messenger" (pay a new cost to forget). Anticipating that others will behave in this way, in turn, allows everyone to more confidently invest in denial at $t = 0$. To avoid this deleterious outcome, organizations and societies will find it desirable to set up *ex-ante guarantees* such as whistle-blower protections, devil's advocates, constitutional rights to free speech, independence of the press, etc. These will ensure that bad news will most likely "resurface" ex-post in a way that is hard to ignore, thus lowering the ex-ante return of investing in denial.

Similar results apply if the dissenter comes at an interim stage, after people have censored but before investments are made. For $s < s^*$ they should welcome the opportunity to correct course, but in practice this can be hard to achieve, requiring full coordination. With payoff heterogeneity, dissenters' motives may also be suspect. Things are even starker for $s > s^*$, meaning that people strongly value hope and dislike anxiety. Facing the truth (state L) now lowers everyone's utility, generating a *universal unwillingness to listen* –the curse of Cassandra. Free-speech guarantees, anonymity and similar protections nonetheless *remain desirable ex-ante,* as they avoid welfare losses in state H and, on average, save the organization or society from wasting resources on denial and repression.

5. Market exuberance

5.1. The dynamics of manias and crashes

I now consider delusions in asset markets. To take recent examples, state H may correspond to a "new economy" in which high-tech startups will flourish and their prospects are best assessed using "new metrics"; to a permanent rise in housing values; or to any other positive and lasting shift in fundamentals. Conversely, state L would reflect an inevitable return to "old" economy and valuations, the unsustainability of many adjustable-rate mortgages, no-docs loans and other subprime debt, or the presence of extensive fraud. Investors finding reasons to believe in H even as evidence of L accumulates corresponds to what Shiller (2005) terms *"new-era thinking",* and of which he relates many examples. This section will provide the first analytical model of this phenomenon.[53]

[53] As explained earlier, neither rational bubbles nor informational cascades involve any element of wishful thinking, motivated rationalization or information avoidance. In both cases, all investors act exactly as a benevolent statistician would advise or allow them to.

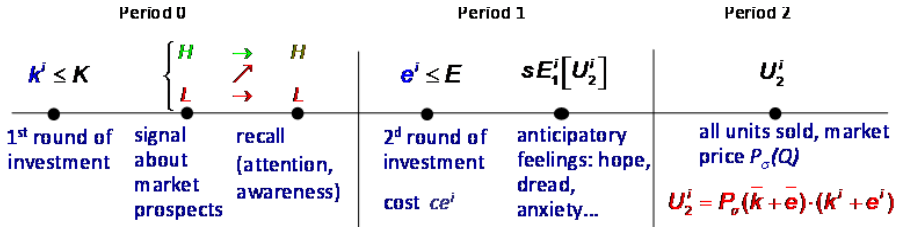

Figure 7: The market game

To this end, I extend the basic framework in two ways, adding an ex-ante investment stage and deriving final payoffs from market prices: see Figure 7.[54] A continuum of firms or investors i can each produce $k^i \le K$ units of a good or asset (housing, office space, mortgage-backed security, internet startup) in period 0 and an additional $e^i \le E$ units in period 1, where K and E reflect capacity constraints or "time to build" technological limits. The cost of production in period 0 is set to 0 for simplicity, while in period 1 it is equal to c. All units are sold at $t = 2$, at which time the expected market price $P_\sigma(Q)$ will reflect total supply $Q \equiv \bar{k} + \bar{e} \in [0, K + E]$ and stochastic market conditions θ_σ, with $\sigma = H, L$ and $P'_\sigma(Q) < 0$. Between the two investment phases agents all observe the signal σ, then decide how to process it, with the same information structure and preferences as before.

The absence of an interim or futures market before date 2 is a version (chosen for simplicity) of the kind of "limits to arbitrage" commonly found in the finance literature. Specifically, I assume that: (i) goods produced in period 0 cannot be sold before period 2, for instance because they are still work-in-progress whose quality or market potential is not verifiable: startup company, unfinished residential development or office complex, new type of financial instrument, etc.; (ii) short sales are not feasible.

Limited liquidity and arbitrage are empirically descriptive of the types of markets which the model aims to analyze.[55] In the recent financial crisis, a dominant fraction of the assets held by major U.S. investment banks did not have an active trading market and objective price, but were instead valued according to the bank's own models and projections, or even according to management's "best estimates".[56] Similarly, the notional value of outstanding

[54] The initial investment stage is an example of endogenizing the degree (previously, $1 - \alpha$) of agents' interdependence or "vesting" in the collective outcome.

[55] Shiller (2003) cites several studies documenting the fact that short sales have never amounted to more than 2% of stocks, whether in number of shares or value. Gabaix et al. (2007) provide specific evidence of limits to arbitrage in the market for mortgage-backed securities.

[56] Reilly (2007) reports that only 36% of Lehman Brothers' 2007-QII balance sheet and 18% of Bear

Collateralized Debt Obligation (CDO) tranches stood in 2008 at about $2 trillion worldwide, and that of Credit Default Swaps (CDS) at around $50 trillion; and yet for most of them there was and still is no established, centralized marketplace where they could easily be traded. These are instead very illiquid ("buy and hold") and *hard-to-price* assets: originating in private deals, highly differentiated and exchanged only over-the-counter.[57]

Suppose that, ex-ante, the market is sufficiently profitable that everyone invests up to capacity at the start of period $0: k^j = \bar{k} = K$.[58] Moreover, following (3), let

$$P_L(K) < \frac{c}{s+\delta} < \frac{c}{\delta} < qP_H(K+E) + (1-q)P_L(K+E). \tag{39}$$

It is thus a dominant strategy for an agent at $t = 1$ to invest the maximum $e^i = E$ if his posterior is no worse than the prior q, and to abstain if he is sure that the state is L.

Consider now what unfolds when agents observe the signal L at the end of period 0.

• *Realism.* If market participants acknowledge and properly respond to bad news ($\lambda^j \equiv 1$) they will not invest further at $t = 1$, so the price at $t = 2$ will be $P_L(K)$. For an individual investor i with stock k^i, the net effect of ignoring the signal is then

$$U^i_{0,D} - U^i_{0,R} = -m + \delta\left[(\delta+s)P_L(K) - c\right]E + \delta s r(\lambda^i)\left[P_H(K+E) - P_L(K)\right](k^i+E). \tag{40}$$

The second term reflects the expected losses from investing at $t = 1$, while the last one represents the value of maintaining hope that the market is strong or will eventually recover, in which case total output will be $K+E$ and the price $P_H(K+E)$. Realism is an equilibrium if $U^i_{0,D} \leq U^i_{0,R}$ for $\lambda^i = 1$ and $k^i = K$, or

$$s \leq \frac{m/\delta + \left[c - \delta P_L(K)\right]E}{\left[P_H(K+E) - P_L(K)\right](K+E) + P_L(K)E} \equiv \underline{s}(1). \tag{41}$$

• *Denial.* If the other participants remain bullish in spite of adverse signals, they will

Stearns' were Level 1 assets in the FASB nomenclature, namely those which "trade in active markets with readily available prices". Level 2 assets ("mark to model") accounted for 56% and 74% respectively, and Level 3 ("reflect management's best estimates of what market participants would use in pricing the assets") for 8% in both cases. For Level 2, moreover, the major trading houses commonly used computer programs designed for "plain vanilla" loans to value novel and highly complex securities (Hansell, (2008)).

[57] In housing, the market for regional-index futures (Case-Shiller) is also still small and fairly illiquid.

[58] The optimality of this first-stage strategy (given expected equilibrium profits in both states) is formally proved in online Appendix C.

keep investing at $t = 1$, *causing the already weak market to crash:* at $t = 2$, the price will fall to $P_L(K + E) < P_L(K)$. The net value of denial for investor i is now

$$U^i_{0,D} - U^i_{0,R} = -m + \delta\left[(\delta + s)P_L(K + E) - c\right]E + \delta s r(\lambda^i)\left[P_H(K + E) - P_L(K + E)\right](k^i + E). \tag{42}$$

In the second term, the expected losses from overinvestment are higher than when other participants are realists. Through this channel, which reflects the usual *substitutability* of investments in a market interaction, each individual's cost of delusion increases when others are deluded. On the other hand, the third term makes clear that the psychological value of denial is also greater, since acknowledging the bad state now requires *recognizing an even greater capital loss* on preexisting holdings. This is again the MAD principle at work.

Denial is an equilibrium if $U^i_{0,D} \geq U^i_{0,R}$ for $\lambda^i = 0$ and $k^i = K$, or

$$s \geq \frac{m/\delta + \left[c - \delta P_L(K + E)\right]E}{q\left[P_H(K + E) - P_L(K + E)\right](K + E) + P_L(K + E)E} \equiv \bar{s}(0). \tag{43}$$

In such an equilibrium, each investor keeps optimistically accumulating assets that have in fact become "toxic", both to his *own* balance sheet and to the *market* at large.

When does other participants' exuberance make each individual more likely to also be exuberant? Intuitively, contagion occurs when the substitutability effect, which bears on the *marginal* units E produced in period 1, is dominated by the capital-loss effect on the *outstanding position* K inherited from period 0. Formally, $\bar{s}(0) < \underline{s}(1)$ requires that K be large enough relative to E, though not so large as to preclude (41).

Proposition 10. ***(Market manias and crashes)*** *If*

$$P_H(K + E)\left(1 + E/K\right) < c/\delta < P_H(K + E), \tag{44}$$

there exists $q^* < 1$ such that, for all $q \in [q^*, 1]$, there is a non-empty interval for s (or c) in which both realism and evidence-blind "exuberance" are equilibria, provided m is not too large. Contagious exuberance leads to overinvestment, followed by a deep crash.

The model provides a microfounded and psychologically-based account of market group-

think, investment frenzies and ensuing crashes.[59] It also identifies key features of the markets prone to such cycles, distinguishing it from traditional models of bubbles or herding.

First, there must be a "story" about shifts in fundamentals that is minimally plausible a priori (q must not be too low): technology, demographics, globalization, etc. The key result is that investors's beliefs in the story can then quickly become resistant to any contrary evidence.[60] Second, when the new opportunity first appears (q rising above the threshold), there is an initial phase of investment buildup and rising price expectations.[61] Finally, the assets in question must involve both significant uncertainty and limited liquidity, as discussed earlier. These conditions are typical of assets tied to new technologies or financial instruments, whose potential will take a long time to be fully revealed.

The model's comparative statics also shed light on other puzzles. From (40)-(43), we have:

(a) *Escalating commitment* at the individual level: the more an agent has invested to date, the more likely he is to continue in spite of bad news, thus displaying a form of the *sunk cost fallacy:* by (42), $\partial(U^i_{0,D} - U^i_{0,R})/\partial k^i > 0$. Moreover, while k^i represents here an outstanding inventory or financial position, any other illiquid asset with market-dependent value, such as sector-specific human capital in banking or finance, has the same effect.[62]

(b) *Market momentum:* the larger the market buildup ($k^{-i} = K$), the more likely is each agent to continue investing in spite of bad news, if demand is (sufficiently) less price sensitive in the low state than in the high one. Indeed, the incentive to discount bad news rises with prospective capital losses, which in a denial equilibrium are proportional to $P_H(K+E) - P_L(K+E)$ and therefore increasing in K when $\partial^2 P/\partial Q \partial \theta > 0$. This occurs for instance with linear demand $Q(P,\theta) = \theta\,(a - bP)$, or when demand is concave and good fundamentals correspond to a scarcity of a close substitute: $P_\sigma(Q) = \mathcal{P}(Q + Z(\theta_\sigma))$, with $Z', \mathcal{P}', \mathcal{P}'' < 0$.[63]

[59] As always, equilibrium multiplicity represents more broadly the potential to greatly amplify small shocks, translating here into a "fragility" of the market to recurrent manias.

[60] By contrast, in standard models of stochastic bubbles everyone realizes they are trading a "hot potato" whose value does not reflect any fundamentals, must eventually collapse and can do so at any instant. Limited liquidity also plays no role there, nor does it in models of herding.

[61] In the interim period there is no objective market price, but all participants' "mark to model" or "best estimates" values remain at $qP_H(K+E)+(1-q)P_L(K+E)$, which reflects only the increased prior q instead of falling to the very low $P_L(K+E)$ actually warranted by the red flags which they are ignoring ($\sigma = L$). Note also that the most economically important aspect of market manias is not price volatility or mispricing per se but the resulting misallocation of resources, which is what the present analysis focuses on.

[62] An initial stake raises the propensity to wishful exuberance, but is not a precondition. Equation (40) or (42) can be positive (for $\lambda^i = 0$) even with $k^i = 0$, given a sufficient sensitivity to anticipatory feelings, s^i.

[63] By (42), $\partial(U^i_{0,D} - U^i_{0,R})/\partial K\big|_{e^j=E} > 0$ at $r(\lambda^i) = q$, so that agent i's best response is $\lambda^i = 0$ (and $e^i = E$),

This simple asset-market model could be extended in several ways. First, in a dynamic context, outstanding stocks will result stochastically from the combination of previous investment decisions and demand realizations. Second, one could relax the strong form of limits to arbitrage imposed by the assumption that trades occur only at $t = 2$. Forward or short trades could instead involve transactions costs, risk due to limited market liquidity or, for large positions, an adverse price impact.[64] Finally, rather than ignoring red flags, the contagion analysis could be recast (as in Section 3) in terms of market participants' unwillingness to seriously examine the true nature –investment-grade, or highly "toxic"– of the assets being accumulated.

5.2. Regulators, politicians and economists

Another set of actors with "value at risk" in an exuberant market are politicians and regulators, whose reputation and career will suffer if the disaster scenario (state L, worsened by market participants' overinvestment) occurs. This should normally make them try to dampen the market's enthusiasm, but if the buildup has proceeded far enough (high K) that large, economy-wide losses are unavoidable in the bad state, they will also become "believers" in a rosy future or smooth landing. Consequently, they will fail to take measures that could have limited (tough not avoided) the damage, thus further enabling the investment frenzy and subsequent crash.[65] Some academics and policy advisers may also have intellectual capital vested in the virtues of unfettered financial markets: a severe crisis proving such faith to be excessive would damage its value and the general credibility of laissez-faire arguments, increasing demand for regulation in other parts of the economy.

---

if and only if $\left[P'_H(K+E) - P'_L(K+E)\right] / \left[-P'_L(K+E)\right] > \left[(\delta+s)/sq\right]\left[E/(k^i+E)\right]$. This inequality holds if $\partial^2 P/\partial Q \partial\theta$ is large enough and k^i/E (equal to K/E in equilibrium) high enough that the right-hand side is less than 1. With linear demand, it becomes $(\theta_H - \theta_L)/\theta_H > \left[(\delta+s)/sq\right]\left[E/(k^i+E)\right]$.

[64] Trying to sell (or sell short) in period 1 could also be self-defeating, as it would reveal again to the market that the state is L, generating an immediate price collapse. For a model of how market thinness generates endogenous limits to arbitrage and delays in trade, see Rostek and Weretka (2008).

[65] On serial blindness to red flags and deliberate information-avoidance by former FED chairman Alan Greenspan and other top financial regulators, see Goodman (2008), SEC (2008, 2009) and Appendix D.

6. Conclusion

This paper developed a model of how wishful thinking and reality denial spread through organizations and markets. The underlying mechanism does not rely on complementarities in technology or preferences, agents' herding on a subset of private signals, or exogenous biases in inference. It is also quite robust to alternative ways of modeling the psychological motives and cognitive operations underlying individual belief distortion, as well as agents' degree of sophistication. This "Mutually Assured Delusion" principle is broadly applicable, helping to explain corporate cultures characterized by dysfunctional groupthink or valuable group morale, why willful ignorance and delusions flow down hierarchies, and the emergence of market manias sustained by "new-era" thinking, followed by deep crashes.

In each of these applications, the institutional and market environment was kept simple, so as to make clear the workings of the underlying mechanism. Enriching these context-specific features would be valuable and permit new applications. For hierarchical organizations, richer payoff and information structures could be incorporated, along with greater heterogeneity of interests among agents. Potential applications include the spread of organizational corruption (e.g., Anand et al. (2005)), corporate politics (e.g. Zald and Berger (1998)) and organizational-design questions such as the optimal mix of agents, network structure and communication mechanisms (e.g. Calvó-Armengol et al. (2009), Van den Steen (2010)). For financial institutions, one could examine how different contractual and regulatory structures may create complementarities in their willingness to find out, or avoid finding out, the true quality of the assets on their balance sheets.

A somewhat different class of collective delusions are mass panics and hysterias. While the model does generate episodes of excessive doubt, overcautiousness and even fatalistic apathy (see online Appendix B), these seem too mild to capture what goes on in a full-fledged panic.[66] Understanding the sources and transmission mechanisms that underlie delusional group pessimism, rather than optimism, is an interesting question for further research.

[66] Recall first that when agents censor bad news, they never fully believe in the good state ($\sigma = H$), even when it actually occurs ($r(\lambda^i) < 1$ for any $\chi > 0$). Second, investors who fear (perhaps from having been burned once) falling prey to the next wave of collective overoptimism will shy away from even positive expected-value investments (this occurs when condition (A.24) in the appendix is reversed).

Appendix A: Main Proofs

In the proofs given here, I maintain the text's focus on cognitive decisions in state L, fixing everyone's recall strategy in state H to $\lambda_H = 1$. In online Appendix C (Lemmas 5 and 6), I show that this is not a binding restriction: with the payoffs (1) there is no equilibrium with $\lambda_H < 1$ and no profitable individual deviation to $\lambda_H^i < 1$ from an equilibrium with $\lambda_H = 1$.[67] These results, as well as Proposition 12, are proved using the more general specification

$$U_2^i \equiv \theta \left[\alpha e^i + (1-\alpha)e^{-i}\right] + \gamma, \tag{A.1}$$

where γ, like θ, is now also state-dependent and $\Delta\gamma \equiv \gamma_H - \gamma_L$ can be of either sign.

Proof of Proposition 1. Parts (ii) and (iii) follow from the monotonicity of Ψ in θ_L and α. Note that no assumption of symmetry in strategies was imposed (λ^{-i} could, a priori, be the mean of heterogenous recall rates). Therefore, the only equilibria are the symmetric ones described in the proposition. ■

Proof of Proposition 2. By Lemma 1, $\lambda = 1$ is an equilibrium when $s \leq \underline{s}(1)$, or $\Psi(1, s|1) \leq 0$ and $\lambda = 0$ is an equilibrium when $s \geq \bar{s}(0)$, or $\Psi(0, s|0) \geq 0$. Finally, $\lambda \in (0,1)$ is an equilibrium if and only if $\Psi(\lambda, s|\lambda) = 0$. Now, from (9) and (6),

$$\Psi(\lambda, s|\lambda) = -m/\delta - c + (\delta + s)\,\alpha\theta_L + sq\left(\frac{\Delta\theta + (1-\alpha)\lambda\theta_L}{q + (1-q)(1-\lambda)}\right). \tag{A.2}$$

This function is either increasing or decreasing in λ, depending on the sign of $(1-\alpha)\theta_L + (1-q)\,\Delta\theta$. One can also check, using (10)-(11), that the same expression governs the sign of $\underline{s}(1) - \bar{s}(0)$. The equilibrium set is therefore determined as follows:

(a) If (14) does not hold, $\Psi(\lambda, s|\lambda)$ is increasing, so $\Psi(0, s|0) < \Psi(1, s|1)$, or equivalently $\underline{s}(1) < \bar{s}(0)$ by (10)-(11). There is then a unique equilibrium, equal to $\lambda = 1$ if $\Psi(1, s|1) \leq 0$, interior if $\Psi(0, s|0) < 0 < \Psi(1, s|1)$, and equal to $\lambda = 0$ if $0 < \Psi(0, s|0)$.

(b) If (14) does hold, $\Psi(\lambda, s|\lambda)$ is decreasing, so $\Psi(1, s|1) < \Psi(0, s|0)$, or equivalently $\bar{s}(0) < \underline{s}(1)$ by (10)-(11). Then: (i) $\lambda = 1$ is the unique equilibrium for $\Psi(0, s|0) \leq 0$, meaning that $s \leq \bar{s}(0)$, while $\lambda = 0$ is the unique equilibrium for $\Psi(1, s|1) \geq 0$, meaning

[67] Under the very weak condition that each agent encodes his own information (for future recall) in a cost-effective manner, which Lemma 5 shows can always be ensured. This is seen most clearly for $\lambda_H^i = \lambda_L^i = 0$, which is informationally equivalent to $\lambda_H^i = \lambda_L^i = 1$ but wastes m in each state.

that $s \geq \underline{s}(1)$; for $\Psi(1,s|1) < 0 < \Psi(0,s|0)$, or $\bar{s}(0) < s < \underline{s}(1)$, both $\lambda = 1$ and $\lambda = 0$ are equilibria, together with the unique solution to $\Psi(\lambda, s|\lambda) = 0$, which is interior. ■

Corollary 1. *Denote by $\underline{s}(\lambda^{-i}, \alpha)$ and $\bar{s}(\lambda^{-i}, \alpha)$ the thresholds respectively given by (10) and (11), and by $\tilde{s} \equiv \underline{s}(\lambda^{-i}, 1)$, which is independent of λ^i. Let $\alpha' < 1$ be such that $\delta[\alpha'\theta_L + \theta_H] > c$ and (14) holds. Then, for all m small enough, $\bar{s}(0,\alpha') < \underline{s}(1,\alpha') < \tilde{s}$ and:*
(i) For $\underline{s}(1,\alpha') < s < \tilde{s}$, $\lambda = 1$ is the unique equilibrium when $\alpha = 1$, and $\lambda = 0$ the unique equilibrium when when $\alpha = \alpha'$;
(ii) For $\bar{s}(0,\alpha') < s < \underline{s}(1,\alpha)$, $\lambda = 1$ is the unique equilibrium when $\alpha = 1$, and $\{0,1\}$ is the stable equilibrium set when $\alpha = \alpha'$.

Proof. The fact that $\bar{s}(0,\alpha') < \underline{s}(1,\alpha')$ is simply equation (14), while $\underline{s}(1,\alpha') < \tilde{s}$ if

$$[m/\delta + c - \delta\alpha'\theta_L][\alpha'\theta_L + \Delta\theta] < [m/\delta + c - \delta\theta_L][\alpha'\theta_L + \Delta\theta + (1-\alpha')\theta_L] \tag{A.3}$$

For $m = 0$, this becomes:

$$\begin{aligned} (c - \delta\alpha'\theta_L)(\alpha'\theta_L + \Delta\theta) &< (\alpha'\theta_L + \Delta\theta + (1-\alpha')\theta_L)(c - \delta\theta_L) \iff \\ (1-\alpha')\delta\theta_L[\alpha'\theta_L + \Delta\theta] &< (1-\alpha')\theta_L(c - \delta\theta_L) \iff \delta[\alpha'\theta_L + \Delta\theta] > c - \delta\theta, \end{aligned} \tag{A.4}$$

since $\theta_L < 0$, by (14). Therefore, since $\delta[\alpha'\theta_L + \theta_H] > c$, (A.3) holds for m small enough. With $\alpha = 1$, the uniqueness of equilibrium follows from $s < \tilde{s} = \underline{s}(1,1)$ and Proposition 2.2. With $\alpha = \alpha'$, results (i) and (ii) respectively follow from parts 2 and 1 of Proposition 2. ■

Proof of Proposition 3. Setting $\lambda^j \equiv 1$ in (18) and $\lambda^j \equiv 0$ in (19) yields the result. ■

Adaptive-learning version of the model. Let the game summarized by Figure 1 be repeated many times, and index those where state L occurs by $\tau \in \mathbb{N}$. At any stage $t = 1$, agent i's optimal decision depends only on his own belief about θ. At stage $t = 0$, by (1)-(2) the only aspect of other agents' play affecting his future payoffs is the aggregate action e_τ^{-i} they will choose at $t = 1$, impacting him by $(1-\alpha)\theta e_\tau^{-i}$. Instead of forecasting e_τ^{-i} by using as before the equilibrium cognitive response λ_τ^{-i} to $\sigma = L$, let each agent now simply best-respond to the aggregate investment level $e_{\tau-1}^{-i}$ observed in the previous (similar) round.[68] For simplicity, and without loss of generality, assume also that:

[68] The state σ drawn in any repetition of the stage game is also assumed to be observable ex-post (at stage $t = 2$, when material payoffs are realized), even by those who temporarily forgot it. Such ex-post observability is in any case irrelevant for full groupthink ($\lambda^j \equiv 0$), where everyone invests in both states.

(i) Consistently with the idea of bounded rationality, agents are unsophisticated about their own cognitive processes, as they are with respect to those of others: $\chi = 0$ in (6);

(ii) Agents form a continuum, with parameter s distributed according to $F(s)$ on $[s_{\min}, s_{\max}]$; heterogeneity could also be with respect to c or θ_H, or idiosyncratic signals about these variables. The continuum assumption will "smooth out" best responses and also equate e_τ^{-i} with the aggregate response (including i's), denoted e_τ.

With agents thus *best responding to past play,* the optimal choice between $\hat{\sigma}_\tau^i = L$ and $\hat{\sigma}_\tau^i = H$ is still governed by comparing (7) and (8), but with $(1-\lambda^{-i})\theta_L$ replaced by $e_{\tau-1}\theta_L$; in addition, $r(\lambda^i)$ simply becomes 1, since $\chi = 0$. The set of realists at any stage $\tau \geq 1$ of this adaptive process therefore consists of the agents with $s^i \leq \underline{s}(1-e_{\tau-1})$, where the function $\underline{s}(\cdot)$ is still given by (10); their proportion is thus $\lambda_\tau = F\left[\underline{s}(1-e_{\tau-1})\right]$. Since realists choose $e_\tau^i = 0$ and deniers $e_\tau^i = 1$, moreover, we have $e_\tau = 1 - \lambda_\tau$. Hence the law of motion

$$\lambda_\tau = F\left(\frac{m/\delta + c - \delta\alpha\theta_L}{\alpha\theta_L + \Delta\theta + (1-\alpha)\lambda_{\tau-1}\theta_L}\right), \quad \forall \tau > 1. \tag{A.5}$$

For $\theta_L > 0$, λ_τ is decreasing in λ_{-1}, generating stable cobweb dynamics converging to a unique equilibrium (steady-state), and a multiplier less than 1 for responses to any local change in parameters. By contrast, when $\theta_L < 0$ the transition function is increasing, generating monotone dynamics, a scope for multiple equilibria (reached from different initial conditions e_0) and a multiplier locally greater than 1 (and increasing in $-\theta_L$).[69] ■

Proof of Lemma 2 and Propositions 5-6. From (29), we have

$$\varphi'(d) \equiv -(1-q)\left[v'\left(qA_H^i + (1-q)\left(B_L^i - f_L^i - d\right)\right) - v'(B_L^i - d)\right], \tag{A.6}$$

so $\varphi'(d) > 0$ if and only if $B_L^i - d < qA_H^i + (1-q)(B_L^i - f_L^i - d)$, or $d > d_*^i$ defined in (32). Therefore, $\varphi(d)$ is strictly quasiconvex, with a minimum at d_*^i. Moreover, $qA_H^i + (1-q)(B_L^i - f_L^i - d_*^i) = B_L^i - d_*^i$, implying $\varphi(d_*^i) = v(B_L^i - d_*^i) - qv(A_H^i) - (1-q)v(B_L^i - d_*^i)$, or

$$\varphi(d_*^i) = q\left[v\left(B_L^i - d_*^i\right) - v(A_H^i)\right] = q\left[v(A_H^i - f_L^i(1-q)/q) - v\left(A_H^i\right)\right] < 0. \tag{A.7}$$

[69] Thus, $\lambda = 1$ and $\lambda = 0$ are both equilibria when $[s_{\min}, s_{\max}] \subset [\underline{s}(0), \underline{s}(1)]$, which can be ensured only when $\theta_L < 0$. There is even a continuum of equilibria for $[s_{\min}, s_{\max}] \equiv [\underline{s}(0), \underline{s}(1)]$ and $F(s) \equiv (\underline{s})^{-1}(s)$. Even with a unique equilibrium (or selecting the one reached from $e_0 = 1$), the multiplier can be made arbitrarily large by appropriate choice of θ_L. Finally, in the limit where F degenerates to a mass-point (homogenous agents), the fixed points of (A.5) coincide exactly with the equilibrium set of Proposition 2 (for $\chi = 0$).

(2) As d tends to $+\infty$, $\varphi^i(d) \approx v(-d(1-q))) - (1-q)v(-d)$, which behaves as $[(1-q) - (1-q)^\gamma] \times (-d)^\gamma$ and thus tends to $+\infty$, since $\gamma' > 1$. Similarly, as d tends to $-\infty$, $\varphi^i(d) \approx v(-d(1-q))) - (1-q)v(-d)$, which behaves as $[(1-q) - (1-q)]^{1/\gamma} \times (d)^{1/\gamma}$and thus tends to $+\infty$, since $1/\gamma < 1$. The rest of Lemma 2 and Proposition 5 follow immediately, as does Proposition 6 since (31) implies $\varphi^i(0) < 0$, hence $\underline{d}^i < 0 < \bar{d}^i$. ■

Proposition 11. *Let $v(x) \equiv x - \gamma x^2/2$, and let (23), (24) and (31) hold for all i. If $|d_L^i|$ is large enough, for all i, there is a non-empty range $[\underline{\gamma}, \bar{\gamma}]$ such the informed uniformed equilibria coexist if and only if $\gamma \in [\underline{\gamma}, \bar{\gamma}]$.*

Proof. Condition (29) takes the form

$$qA_H^i + (1-q)\left(B_L^i - f_L^i - d_L^i\right) - (\gamma/2)\left[qA_H^i + (1-q)\left(B_L^i - f_L^i - d_L^i\right)^2\right] >$$
$$qA_H^i + (1-q)\left(B_L^i - d_L^i\right) - (\gamma/2)\left[q\left(A_H^i\right)^2 + (1-q)\left(B_L^i - d_L^i\right)^2\right] \iff$$
$$(1-q)f_L^i + (\gamma/2)\left[qA_H^i + (1-q)\left(B_L^i - f_L^i - d_L^i\right)\right]^2 < (\gamma/2)\left[q\left(A_H^i\right)^2 + (1-q)\left(B_L^i - d_L^i\right)^2\right],$$

which is equivalent to (33). Similarly, (30) is equivalent to (34). Together, they define a nonempty range for γ if and only if

$$q\left(A_H^i - B_L^i + f_L^i\right)^2 - f_L^i\left(f_L^i - 2B_L^i\right) < q\left(A_H^i - B_L^i + d_L^i + f_L^i\right)^2 - f_L^i\left(f_L^i - 2B_L^i + 2d_L^i\right) \iff$$
$$2f_L^i d_L^i < q\left(\left(d_L^i\right)^2 + 2d_L^i\left(A_H^i - B_L^i + f_L^i\right)\right).$$

If $d_L^i > 0$, which is the main case of interest, this inequality becomes:

$$d_L^i > 2\left[(1-q)f_L^i/q - \left(A_H^i - B_L^i\right)\right] = 2d_*^i. \tag{A.8}$$

which holds for d large enough –e.g., for all $d > 0$ when $d_*^i < 0$. If $d_L^i < 0$, the condition is reversed, and thus holds for $-d_L^i$ large enough (in e.g., for all $d < 0$ when $d_*^i > 0$).

Recalling finally that the highest possible payoff, A_H^i, must lie in the interval $(-\infty, 1/\gamma)$ over which $v(x) = x - \gamma x^2/2$ is increasing, it must also be that $\bar{\gamma}A_H^i < 1$, or

$$\begin{aligned} 2f_L^i A_i^H + f_L^i\left(f_L^i - 2B_L^i + 2d_L^i\right) &< q\left(A_H^i - B_L^i + d_L^i + f_L^i\right)^2 \iff \\ 2q\left(A_H^i - B_L^i + d_L^i\right)^2 &> 2(1-q)f_L^i\left(A_H^i - B_L^i + d_L^i\right) + (1-q)\left(f_L^i\right)^2. \end{aligned} \tag{A.9}$$

Define the polynomial $P(X) \equiv qX^2 - 2(1-q)f_L^i X - (1-q)\left(f_L^i\right)^2$. The discriminant is

$\Delta' = (1-q)\left(f_L^i\right)^2$, therefore the required condition is

$$\text{(A.10)} \qquad \left(q/f_L^i\right)\left(A_H^i - B_L^i + d_L^i\right) \notin \left((1-q) - \sqrt{1-q},\ (1-q) + \sqrt{1-q}\right),$$

which again holds if $|d_L^i|$ is sufficiently large. ■

Proof of Proposition 8. Part (1) follows directly from (36) and (12)-(13). In Part (2), it is easily seen that $s^* < \bar{s}(0)$, but $s^* < \underline{s}(1)$ requires $(1-q)\,\Delta\theta[m/\delta + c - \delta\alpha\theta_L] < \delta\,(1-\alpha)\,\theta_L\theta_H$, which can go either way.

Proof of Proposition 10. Assume for now that at $t=0$, everyone else invests $k^{-i} = K$. Since investing (respectively, abstaining) at $t=1$ is a dominant strategy given posterior $\mu^j = r(\lambda^j) \geq q$ (respectively, $\mu^j = 0$), the price in state L will be $P_L(K + (1-\lambda^{-i})E)$ and the date-0 expected utilities of realism and denial equal to

$$\text{(A.11)} \quad U_{L,R}(\lambda^i, \lambda^{-i}; k^i)/\delta = (\delta + s)P_L(K + (1-\lambda^{-i})E)k^i,$$

$$\begin{aligned}\text{(A.12)} \quad U_{L,D}(\lambda^i, \lambda^{-i}; k^i_{,})/\delta &= -m/\delta + (\delta+s)P_L(K + (1-\lambda^{-i})E)(k^i + E) - cE \\ &\quad + sr(\lambda^i)\left[P_H(K+E) - P_L(K + (1-\lambda^{-i})E)\right](k^i + E).\end{aligned}$$

The net incentive for denial, $\Delta U_L \equiv U_{L,D} - U_{L,R}$, is thus given by

$$\begin{aligned}\text{(A.13)} \quad [\Delta U_L(\lambda^i, \lambda^{-i}; \bar{k};) + m]/\delta &= \left[(\delta+s)P_L(K + (1-\lambda^{-i})E) - c\right]E, \\ &\quad + sr(\lambda^i)\left[P_H(K+E) - P_L(K + (1-\lambda^{-i})E)\right](k^i + E).\end{aligned}$$

Setting $r(\lambda^i) = 1$, realism is a (personal-equilbrium) best response to λ^{-i} for an agent entering period 1 with stock k^i if

$$\begin{aligned}\text{(A.14)} \qquad m/\delta &\geq \left[(\delta+s)P_L(K + (1-\lambda^{-i})E) - c\right]E \\ &\quad + s\left[P_H(K+E) - P_L(K + (1-\lambda^{-i})E)\right](k^i + E).\end{aligned}$$

Conversely, denial ($r(\lambda^i) = q$) is a (personal-equilibrium) best response for i if

$$\begin{aligned}\text{(A.15)} \qquad m/\delta &\leq \left[(\delta+s)P_L(K + (1-\lambda^{-i})E) - c\right]E \\ &\quad + sq\left[P_H(K+E) - P_L(K + (1-\lambda^{-i})E)\right](k^i + E).\end{aligned}$$

For given k^i and λ^{-i}, these two conditions are mutually exclusive. When neither holds, there is a unique $\lambda^i \in (0,1)$ that equates ΔU_L to zero, defining a mixed-strategy (personal

equilibrium) best-response. The next step is to solve for (symmetric) social equilibria.

1. Realism. From (A.14), $\lambda^i = \lambda^{-i} = 1$ is an equilibrium in cognitive strategies if

$$[(\delta + s)P_L(K) - c]\,E + s\,[P_H(K+E) - P_L(K)]\,(k^i + E) \leq m/\delta. \tag{A.16}$$

This condition holds for all $k^i \leq K$ if and only if

$$s \leq \frac{m/\delta + [c - \delta P_L(K)]\,E}{[P_H(K+E) - P_L(K)]\,(K+E) + P_L(K)E} \equiv \underline{s}(1;K). \tag{A.17}$$

Moving back to the start of period 0, one now verifies that it is indeed an equilibrium for everyone to invest $k^i = K$. Since agents will respond to market signals $\sigma = H, L$, the expected price is $qP_H(K+E) + (1-q)P_L(K) > 0$, whereas the cost of period-0 production is 0 (more generally, sufficiently small). Thus, it is optimal to produce to capacity.

2. Denial From (A.15), $\lambda^i = \lambda^{-i} = 0$ is a cognitive equilibrium if

$$[(\delta + s)P_L(K+E) - c]\,E + sq\,[P_H(K+E) - P_L(K+E)]\,(k^i + E) \geq m/\delta. \tag{A.18}$$

This condition holds for $k^i = K$ if

$$s > \frac{m/\delta + [c - \delta P_L(K+E)]\,E}{q\,[P_H(K+E) - P_L(K+E)]\,(K+E) + P_L(K+E)E} \equiv \bar{s}(0;q,K). \tag{A.19}$$

An agent with low k^i, however, has less incentive to engage in denial. In particular, for $s < \underline{s}(1;K)$, (A.16) for $k^i = 0$ precludes (A.18) from holding at $k^i = 0$. Let $\bar{k}(s,q)$ therefore denote the unique solution in k^i to the linear equation

$$[(\delta + s)P_L(K+E) - c]\,E + sq\,[P_H(K+E) - P_L(K+E)]\,(k^i + E) = m/\delta. \tag{A.20}$$

Subtracting the equality obtained by evaluating (A.18) at $s = \bar{s}(0;q,K)$ yields

$$\begin{aligned} & sq\,[P_H(K+E) - P_L(K+E)]\,(K - \bar{k}) \\ = \;& (s - \bar{s})P_L(K+E)E + (s - \bar{s})q\,[P_H(K+E) - P_L(K+E)]\,(K+E), \end{aligned}$$

where the arguments are dropped from $\bar{k}$ and $\bar{s}$ when no confusion results. Thus,

$$K - \bar{k} = \frac{s - \bar{s}}{s} \times \left(\frac{qP_H(K+E) + (1-q)P_L(K+E)}{q\,[P_H(K+E) - P_L(K+E)]} E + K \right) > \frac{s - \bar{s}}{s} \times (K+E)\,. \tag{A.21}$$

Note that $\bar{k} \leq K$ (and is thus feasible) if and only if $s \geq \bar{s}$. One can now examine the optimal choice of k^i at $t = 0$, which will be either $k^i = K$ or some $k^i \leq \bar{k}$.

(a) For $k^i > \bar{k}(s,q)$, (A.20) implies that denial is the unique best response to $\lambda^{-i} = 0$, leading agent i to produce $e^i = E$ in both states at $t = 1$. These units and the initial k^i will be sold at the expected price $\bar{P}_q(K+E) \equiv qP_H(K+E) + (1-q)P_L(K+E) > 0$. Therefore, producing K in period 0 is optimal among all levels $k^i > \bar{k}(s,q)$, and yields ex-ante utility

$$U_D(0,K,K)/\delta = (\delta+s)\bar{P}_q(K+E)(K+E) - cE - (1-q)m/\delta. \tag{A.22}$$

(b) For $k^i \leq \bar{k}(q;s)$, on the other hand, agent i's continuation (personal-equilibrium) strategy is some $\lambda^i = \lambda(k^i) \geq 0$: in state L he weakly prefers to be a realist, achieving

$$\begin{aligned} U(\lambda^i, 0, k^i_{,}K)/\delta &= (\delta+s)\bar{P}_q(K+E)\left(k^i+E\right) - cE \\ &\quad - (1-q)\left\{\left(1-\lambda^i\right)m/\delta - \lambda^i\left[c-(\delta+s)P_L(K+E)\right]E\right\}. \end{aligned} \tag{A.23}$$

The agent prefers $k^i = K$ to any $k^i \leq \bar{k}(q;s)$ if $U_D(0,K,K) > U(\lambda^i, 0, k^i_{,}K)$, or

$$(\delta+s)\bar{P}_q(K+E)(K-k^i) > (1-q)\,\lambda^i\left\{m/\delta + \left[c-(\delta+s)P_L(K+E)\right]E\right\}. \tag{A.24}$$

Using (A.21) and $\lambda^i \leq 1$, it suffices that

$$\left(\frac{s-\bar{s}(0;q,K)}{s}\right)\left(\frac{\bar{P}_q(K+E)\,(K+E)}{1-q}\right) \geq \frac{m}{\delta\,(\delta+s)} + \left(\frac{c}{\delta+s} - P_L(K+E)\right)E. \tag{A.25}$$

Since $\bar{P}_q(K+E)$ tends to $P_H(K+E)$ as q tends to 1, (A.25) will hold for q close enough to 1, provided $s - \bar{s}(0;q,K)$ remains bounded away from 0. Lemmas 3 and 4 (in online Appendix C) formalize this idea, showing that there exist a threshold $q^*(K) < 1$ and a nonempty interval $S^*(K)$ such that, for all $q > q^*(K)$: $S^*(K) \subset (\bar{s}(0;q,K), \underline{s}(1;K))$ and (A.25) holds for all $s \in S^*(K)$. Consequently, when $q > q^*(K)$ both $(k^i = K, \lambda^i = 1)$ and $(k^i = K, \lambda^i = 0)$ are equilibria of the two-stage market game, for any $s \in S^*(K)$. Indeed, we showed that: (i) for $s < \underline{s}(1;K)$, when others play $(k^{-i} = K,\ \lambda^{-i} = 1)$ agent i finds it optimal to also invest $k^i = K$ and then be a realist; (ii) for $s > \bar{s}(0;q,K)$, when others play $(k^{-i} = K,\ \lambda^{-i} = 0)$ he finds it optimal to invest K in period 0 even though he knows that this will cause him to engage in denial if state L occurs. ■

REFERENCES

Akerlof, G., and W. Dickens (1982) "The Economic Consequences of Cognitive Dissonance," *American Economic Review,* 72, 307-19.

Anand, V., Ashforth, B. and J. Mahendra (2005) "Business as Usual: The Acceptance and Perpetuation of Corruption in Organizations," *Academy of Management Executive,* 19, 9-23.

Asch, S. (1956) "Studies of Independence and Conformity: a Minority of One Against a Unanimous Majority. *Psychological Monographs*, 70 (Whole no. 416).

Banerjee, A.(1992) "A Simple Model of Herd Behavior," *Quarterly Journal of Economics,* 107(3), 797-817.

Bazerman, M. and A. Tenbrunsel (2011) *Blind Spots: Why We Fail to Do What's Right and What to Do About It.* Princeton University Press.

Bénabou, R. (2008) "Ideology," *Journal of the European Economic Association,* 6(2), 321–52.

Bénabou, R. and J. Tirole (2002) "Self–Confidence and Personal Motivation," *Quarterly Journal of Economics*, 117 , 871–915.

Bénabou, R. and J. Tirole (2003) ""Intrinsic and Extrinsic Motivation," *Review of Economic Studies,* 70(3), 489-520.

Bénabou, R. and J. Tirole (2004) "Willpower and Personal Rules," *Journal of Political Economy,* 112, 848–887.

Bénabou, R. and J. Tirole (2006a) "Incentives and Prosocial Behavior," *American Economic Review,* 96(5), December , 1652-78.

Bénabou, R. and J. Tirole (2006b) "Belief in a Just World and Redistributive Politics," *Quarterly Journal of Economics,* 121(2), May, 699-746.

Bénabou, R. and J. Tirole (2011) "Identity, Morals and Taboos: Beliefs as Assets," *Quarterly Journal of Economics* 126, 805-855.

Bernheim, D. and R. Thomadsen (2005) "Memory and Anticipation," *The Economic Journal,* 115, 271–304.

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

Brunnermeier, M. and J. Parker (2005) "Optimal Expectations," *American Economic Review,* 90, 1092-118. .

Brunnermeier, M., Gollier, C. and J. Parker (2007) "Optimal Beliefs, Asset Prices, and the Preference for Skewed Returns," *American Economic Review,* 97(2), 159-65.

Calvó-Armengol, A., de Martí, J. and A. Prat (2009) "Endogenous Communication in Complex Organization," LSE mimeo, March.

Camerer, C. and U. Malmendier (2007) "Behavioral Economics of Organizations," in *Behavioral Economics and Its Applications,* P. Diamond and H. Vartiainen (eds.), Princeton University Press.

Caplin, A. and K. Eliaz (2003) "AIDS Policy and Psychology: A Mechanism-Design Approach," *Rand Journal of Economics,* 34(4), 631-646

Caplin, A. and J. Leahy (1994) "Business as Usual, Market Crashes, and Wisdom After the Fact," *American Economic Review,* 84(3), 548-65.

Caplin, A. and J. Leahy (2001) "Psychological Expected Utility Theory and Anticipatory Feelings," *Quarterly Journal of Economics,* 116, 55–79.

Chamley, C. and D. Gale (1994) "Information Revelation and Strategic Delay in a Model of Investment," *Econometrica,* 62(5), 1065-85.

Choi, D. and D. Lou (2010) "A Test of the Self-Serving Attribution Bias: Evidence from Mutual Funds," Hong Kong University of Science and Technology, mimeo, August.

Cialdini, R. (1984) *Influence: The Psychology of Persuasion.* HarperCollins Publishers.

Cohan, J. (2002) " 'I Didn't Know' and 'I Was Only Doing My Job': Has Corporate Governance Careened Out of Control? A Case Study of Enron's Information Myopia". *Journal of Business Ethics,* 40, 275-99.

Columbia Accident Investigation Board (2003) *CIAB Final Report,* especially Chapters 6, 7 and 8. Available at http://caib.nasa.gov/.

Compte, O. and Postlewaite, A. (2004) "Confidence-Enhanced Performance," *American Economic Review,* 94(5), 1536-1557.

Dessi, R. (2008) "Collective Memory, Cultural Transmission and Investments," *American Economic Review,* 98(1), 534-560.

Di Tella, R., Galiani, S., and E. Schargrodsky, (2007) "The Formation of Beliefs: Evidence from the Allocation of Land Titles to Squatters," *Quarterly Journal of Economics,* 122(1), 209-41.

Eichennwald, K. (2005) *Conspiracy of Fools: A True Story. New York, NY:* BroadwayBooks.

Eil, D. and Rao, J. (2011) "The Good News-Bad News Effect: Asymmetric Processing of Objective Information about Yourself," *American Economic Journal: Microeconomics,* 3(2), 114–38.

Eliaz, K. and R. Spiegler (2006) "Can Anticipatory Feelings Explain Anomalous Choices of Information Sources?" *Games and Economic Behavior*, 56 (1), 87-104.

Eyster, E. and Rabin, M. (2009) "Rational and Naive Herding", LSE mimeo, June.

Fang, H., and Moscarini, G. (2005) "Morale Hazard," *Journal of Monetary Economics,* 52(4), 749-78.

Fudenberg, D. and D. Levine (1998) *The Theory of Learning in Games.* Cambridge, MA: MIT Press.

Gabaix, X., Krishnamurthy, A. and O. Vigneron (2007) "Limits of Arbitrage: Theory and Evidence from the Mortgage-Backed Securities Market", *Journal of Finance,* 62(2), 557-595,

Gervais, S. and Goldtsein, I. (2007) "The Positive Effects of Self-Biased Perceptions in Teams," *Review of Finance,* 11(3), 453-96.

Goeree, J., Palfrey, T., Rogers, B. and McKelvey, B. (2007) "Self-Correcting Information Cascades," *Review of Economic Studies,* 74, 733-762.

Goetzman, W. and Peles (1997) "Cognitive Dissonnance and Mutual Fund Investors," *Journal of Financial Research,* 20(2), 145-158.

Goodman, P. (2008) "The Reckoning: Taking Hard New Look at a Greenspan Legacy," *The New York Times,* October 8.

Hansell, S. (2008) "How Wall Street Lied to Its Computers," *The New York Times,* September 18.

Haslam, A. (2004). *Psychology in Organizations: The Social Identity Approach* (2nd ed.). London, UK & Thousand Oaks, CA: Sage.

Hedden, T., Prelec, D., Mijovic-Prelec, D. and J. Gabrieli (2008) "Neural Correlates Of Reward-Related Self-Delusion," Poster Presentation, Cognitive Neuroscience Society Conference, San Francisco, April 2. Available at lawweb.usc.edu/centers/scip/assets/docs/neuro /drazenprelec.ppt.

Hermalin, B. (1998) "An Economic Theory of Leadership: Leading by Example," *The American Economic Review,* 88(5), 1188-206.

Hersh, S. (2004) *Chain of Command.* New York, NY: HarperCollins Publishers.

Huseman, R. and R. Driver (1979) "Groupthink: Implications for Small Group Decision Making in Business," in *Readings in Organizational Behavior: Dimensions of Management Action,* R. Richard Huseman and Archie Carral, eds.. Boston, MA: Allyn and Bacon.

Isikoff, M. and D. Corn (2007) *Hubris.* New York, NY: Three Rivers Press.

Janis, I. (1972) *Victims of Groupthink: Psychological Studies of Policy Decisions and Fiascoes.* Boston, MA: Houghton Mifflin Company.

Karlsson, N., Loewenstein, G. and D. Seppi (2009) "The 'Ostrich Effect': Selective Avoidance of Information," *Journal of Risk and Uncertainty,* 38(2), 95-115.

Kindleberger, C. and R. Aliber (2005) *Manias, Panics, and Crashes: A History of Financial Crises.* Hoboken, NJ: John Wiley and Sons.

Köszegi, B. (2006) "Emotional Agency," *Quarterly Journal of Economics,* 21(1), 121-56.

Köszegi, B. (2010) "Utility from Anticipation and Personal Equilibrium," *Economic Theory,* 44, 415-444.

Kreps, D. and Porteus, E. (1978), "Temporal Resolution of Uncertainty and Dynamic Choice Theory," *Econometrica,* 46(1), 185–200.

Kunda, Z. (1987) "Motivated Inference: Self-Serving Generation and Evaluation of Causal Theories," *Journal of Personality and Social Psychology,* 53(4), 636-647.

Kuran, T. (1993) "The Unthinkable and the Unthought," *Rationality and Society,* 5, 473-505.

Lalancette, M-F. and L. Standing (1990) "Asch Fails Again,", *Social Behavior and Personality,* 18(1),7-12.

Landier, A. (2000) "Wishful Thinking: A Model of Optimal Reality Denial," MIT mimeo.

Landier, A., Sraer, D. and D. Thesmar (2009) "Optimal Dissent in Organizations," *Review of Economic Studies,* 76, 761-794.

Loewenstein, G. (1987) "Anticipation and the Valuation of Delayed Consumption," *Economic Journal*, 97, 666-84.

Mackay, C. (1980) *Extraordinary Popular Delusions and the Madness of Crowds.* New York, NY: Three Rivers Press.

Malmendier, U. and G. Tate (2005) "CEO Overconfidence and Corporate Investment," *Journal of Finance,* 60 (6), 2661-700.

Malmendier, U. and G. Tate (2008) "Who Makes Acquisitions? CEO Overconfidence and the Market's Reaction," *Journal of Financial Economics,* 89(1), 20-43.

Mayraz, G. (2011) "Wishful Thinking," Oxford University Mimeo, October.

Mijovic-Prelec, D. and D. Prelec (2010) "Self-Deception As Self-Signalling: A Model And Experimental Evidence," *Philosophical Transactions of the Royal Society,* B 365, 227–240.

Mischel, W., E. Ebbesen and A. Zeiss (1976) "Determinants of Selective Memory about the Self," *Journal of Consulting and Clinical Psychology,* 44, 92-103.

Möbius, M., Niederle, M., Niehaus, P. and Rosenblat, T. (2010) "Managing Self-Confidence: Theory and Experimental Evidence," Stanford University mimeo, October.

Morgenson, G. and G. Fabrikant (2007) "Countrywide's Chief Salesman and Defender," *The New York Times,* November 2007.

Norris, F. (2008) "Color-Blind Merrill in a Sea of Red Flags." *New York Times,* May 16.

Ottaviani, M. and P. Sørensen (2001)"Information Aggregation In Debate: Who Should Speak First?", *Journal of Public Economics,* 81, 393-421.

Pearlstein, S. (2006) "Years of Self-Deception Killed Enron and Lay," *The Washington Post,* July 8.

Prat, A. (2005) "The Wrong Kind of Transparency," *American Economic Review,* 95(3), 62-877.

Prendergast, C. (1993) "A Theory of 'Yes Men'," *American Economic Review,* 83(4), 757-70.

Reilly, D. (2007) "Marking Down Wall Street." *The Wall Street Journal,* September 14, C1.

Reinhart, C. and Rogoff, K. (2009) *This Time Is Different: Eight Centuries of Financial Folly.* Princeton, NJ: Princeton University Press.

Rogers Commission (1986). *Report of the Presidential Commission on the Space Shuttle Challenger Accident.* http://history.nasa.gov/rogersrep/genindex.htm.

Rostek, M. and M. Weretka (2008) "Dynamic Thin Markets," University of Madison-Wisconsin mimeo, December.

Rotemberg, J. and G. Saloner (2000) "Visionaries, Managers, and Strategic Direction," *Rand Journal of Economics* 31, Winter, 693-716.

Samuelson, R. (2001) "Enron's Creative Obscurity." *The Washington Post,* December 19.

Schelling, T. (1986) The Mind as a Consuming Organ," in D. Bell, Raiffa H. and A. Tversky, eds., *Decision Making : Descriptive, Normative, and Prescriptive Interactions.* Cambridge, MA: Cambridge University Press.

Schrand, C. and S. Zechman (2008) "Executive Overconfidence and the Slippery Slope to

Fraud," Wharton School mimeo, University of Pennsylvania, December.

Securities and Exchange Commission (2008) *SEC's Oversight of Bears Stearns and Related Entities: Consolidated Supervised Entity Program.* Inspector General's Report, Office of Audits, Report No. 446-. September 25, viii-ix. Available at http://www.sec-oig.gov.

Securities and Exchange Commission (2009) *Investigation of Failure of the SEC To Uncover Bernard Madoff's Ponzi Scheme.* Office of Investigations. Case No. OIG-509, August 31. Available at http://www.sec.gov/news/studies/2009/oig-509.pdf.

Shiller, R. (2003) "From Efficient Markets Theory to Behavioral Finance," *Journal of Economic Perspectives,* 17(1), 83-104

Shiller, R. (2005) *Irrational Exuberance.* Second Edition, Princeton University Press.

Sims, R. (1992) "Linking Groupthink to Unethical Behaviors in Organizations," *Journal of Business Ethics,* 11, 651-62.

Slovic, P. (2007) "If I Look at the Mass I will Never Act: Psychic Numbing and Genocide," *Judgment and Decision-Making,* 2(2), 79-95.

Small, D., Loewenstein, G. and Slovic, P. (2007) "Sympathy and Callousness: The Impact of Deliberative Thought on Donations to Identifiable and Statistical Victims," *Organizational Behavior and Human Decision Processes,* 143–53.

Suskind, R. (2004) "Without a Doubt," *The New York Times,* October 17.

Tenbrunsel, A. and D. Messick (2004) "Ethical Fading: The Role of Self-Deception in Unethical Behavior," *Social Justice Research,* 17(2), 223-62.

Van den Steen, E. (2005) "Organizational Beliefs and Managerial Vision," *Journal of Law, Economics and Organization,* 21, 256-283.

Van den Steen, E. (2010) "On the Origins of Shared Beliefs (and Corporate Culture)," *Rand Journal of Economics* 41(4), 617-648.

Von Hippel, W. and R. Trivers (2011) "The Evolution and Psychology of Self-Deception," *Behavioral And Brain Sciences,* 34, 1-56.

Zald, M. and M. Berger (1978) "Social Movements in Organizations: Coup d'Etat, Insurgency, and Mass Movements," *The American Journal of Sociology,* 83(4), 823-861.

Weiszacker, G. (2010) "Do We Follow Others When We Should? A Simple Test of Rational Expectations," *American Economic Review,* 100, 100, 2340-2360.