Confidence collapse in a multihousehold, self-reflexive DSGE model

Federico Guglielmo Morelli[a,b,c], Michael Benzaquen[b,c,d,1], Marco Tarzia[a,e], and Jean-Philippe Bouchaud[c,d,f]

[a]Laboratoire de Physique Théorique de la Matière Condensée, UMR CNRS 7600, Sorbonne Université, 75252 Paris Cedex 05, France; [b]Laboratoire d'Hydrodynamique de l'X, UMR CNRS 7646, Ecole Polytechnique, 91128 Palaiseau Cedex, France; [c]Chair of Econophysics and Complex Systems, Ecole Polytechnique, 91128 Palaiseau Cedex, France; [d]Capital Fund Management, 75007 Paris, France; [e]Institut Universitaire de France, 75231 Paris Cedex 05, France; and [f]Académie des Sciences, 75006 Paris, France

Edited by Jose A. Scheinkman, Columbia University, New York, NY, and approved March 10, 2020 (received for review July 17, 2019)

We investigate a multihousehold dynamic stochastic general equilibrium (DSGE) model in which past aggregate consumption impacts the confidence, and therefore consumption propensity, of individual households. We find that such a minimal setup is extremely rich and leads to a variety of realistic output dynamics: high output with no crises; high output with increased volatility and deep, short-lived recessions; and alternation of high- and low-output states where a relatively mild drop in economic conditions can lead to a temporary confidence collapse and steep decline in economic activity. The crisis probability depends exponentially on the parameters of the model, which means that markets cannot efficiently price the associated risk premium. We conclude by stressing that within our framework, narratives become an important monetary policy tool that can help steer the economy back on track.

DSGE | multihousehold | confidence collapse | economic crises

Despite their poor performance during the global financial crisis (GFC), dynamic stochastic general equilibrium (DSGE) models still constitute the workhorse of monetary policy models around the world (see, e.g., ref. 1 for an insightful introduction and references). Many ingredients that were missing in previous versions of the model (such as the absence of a financial sector) have been added in recent years, in an attempt to assuage some of the scathing criticisms that were uttered post-GFC (see, for example, refs. 2–8 for rebuttals). However, the whole DSGE framework seems to be—partly for technical reasons—wedded to the representative agent (or firm) paradigm and to a (log-)linear approximation scheme that describes small perturbations away from a fundamentally stable stationary state.[†] In other words, crises are difficult to accommodate within the scope of DSGE, a situation that led to J. C. Trichet's infamous complaint: "Models failed to predict the crisis and seemed incapable of explaining what was happening [...], in the face of the crisis we felt abandoned by conventional tools" (ref. 3, p. 3).

Agent-based models (ABMs) provide a promising alternative framework to think about macroeconomic phenomena (9–13). In particular, ABMs easily allow for heterogeneities and interactions. These may generate nonlinear effects and unstable self-reflexive loops that are most likely at the heart of the 2008 crisis, while being absent from benchmark DSGE models where only large technology shocks can lead to large output swings. Unfortunately, ABMs are still in their infancy and struggle to gain traction in academic and institutional quarters [with some major exceptions, such as the Bank of England (14) or the Organisation for Economic Co-operation and Development]. To bridge the gap between DSGE and ABMs and allow interesting nonlinear phenomena, such as trust collapse, to occur within DSGE, we replace the representative household by a collection of homogeneous but interacting households. Interaction here is meant to describe the feedback of past aggregate consumption on the sentiment (or confidence) of individual households—i.e., their future consumption propensity. Low past consumer index begets low future individual consumption (15–17). This opens the possibility that a relatively mild drop in economic conditions leads to a confidence collapse and a steep decline in economic activity.

We establish the "phase diagram" of our extended model and identify regions where crises can occur. As the strength of feedback increases, the economy can undergo rare, short-lived crises, where output and consumption plummet but quickly recover. For even larger feedback, mild technology shocks can induce transitions between a high-output state and a low-output state in which the economy can linger for a long time. In such conditions, output volatility can be much larger than the total factor productivity. As portrayed by Bernanke et al. (18), this is a "small shocks, large business cycle" situation. We show that these endogenous crises exist even when the amplitude of technology shocks is vanishingly small. But in this limit the probability for such crises is exponentially small and hence, we argue, unknowable and unhedgeable. Our model thus provides an interesting example of unknown knowns, where the crisis is a possible state of the world, but its probability is fundamentally uncomputable.

Our work relates to various strands of the literature that emphasize the role of multiple equilibria and self-fulfilling prophecies, in particular the work of Brock and Durlauf (19) on social interactions. Technically, our modeling strategy is akin to the "habit formation" or "keeping up with the Joneses" (KUJ) literature (20–22), although our story is more about keeping

Significance

Despite their inability to cope with the global financial crisis, dynamic stochastic general equilibrium (DSGE) models are still at the forefront of monetary policy. Like many standard economic models, DSGE models rely on the figment of representative agents, abolishing the possibility of genuine collective effects (such as the 2008 crisis) induced by heterogeneities and interactions. By allowing feedback of past aggregate consumption on the sentiment of individual households, we construct more realistic DSGE models that allow for large output swings induced by relatively minor variations in economic conditions and amplified by interactions. Our framework elicits the de facto impossibility to price extreme risks and highlights the importance of narratives, as an efficient depression-prevention policy tool whenever confidence collapse is looming.

Author contributions: M.B., M.T., and J.-P.B. designed research; F.G.M. performed research; and M.B., M.T., and J.-P.B. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: michael.benzaquen@polytechnique.edu.

[†]Quoting O. Blanchard in ref. 4, p. 28: "We in the field did think of the economy as roughly linear, constantly subject to different shocks, constantly fluctuating, but naturally returning to equilibrium over time. [...]. The problem is that we came to believe that this was indeed the way the world worked."

"down" with the Joneses (KDJ), as we are more concerned by self-reflexive confidence collapse than by consumption sprees (which also exist in our model, but in a regime that does not seem to be empirically relevant).

Several other scenarios can lead to the coexistence of static equilibria, corresponding to high/low confidence (23–25), high/low output (11, 26, 27), or high/low inflation expectations (26, 28) or trending/mean-reverting markets (29–32), etc., with possible sudden shifts between the two. Multiple equilibria can be a result of learning either from past events or from strong interactions between individual agents (direct or mediated by markets)—for a review, see ref. 33. Another, distinct line of research explores the consequences of having an indeterminate equilibrium, i.e., a stationary solution around which small fluctuations can develop without being pinned by initial conditions (for a recent review and references, see ref. 34). These fluctuations are not related to any real economic driving force, but rather the result of self-fulfilling prophecies. In our present model, fluctuations are triggered by real technology shocks, but are then amplified by a self-reflexive mechanism. Nothing would prevent, however, the existence of further indeterminacy around different stationary points.

A Multihousehold DSGE Model

We assume that each household $i \in [\![1, M]\!]$ is characterized by a utility function $U_i(c_t^i, n_t^i)$ that depends on its (unique good) consumption c_t^i and amount of labor n_t^i as

$$U_i(c_t^i, n_t^i) = f_t^i \frac{(c_t^i)^{1-\varsigma}}{1-\varsigma} - \gamma_i \frac{(n_t^i)^{1+\phi}}{1+\phi}, \qquad [1]$$

where γ_i is a factor measuring the disutility of labor, and $\varsigma \in \ ]0,1]$ and $\phi > 0$ are two i-independent parameters such that the utility function has the correct concavity. Standard choices are $\varsigma = 1$ (log-utility of consumption) and $\phi = 1$. The quantity f_t^i is a time-dependent factor measuring the confidence of household i at time t and hence their propensity to consume. This "belief function" (34) will be responsible for the possible crises in our model (see below).

Each infinitely lived household maximizes its future expected discounted utility with a discount factor $\beta < 1$, subject to the budget constraint (1)

$$p_t c_t^i + \frac{B_t^i}{1+r_t} \le w_t n_t^i + B_{t-1}^i + E_t^i, \qquad [2]$$

where p_t is the price of the good, w_t the wage (assumed to be identical for all households), and E_t^i any extra source of income (dividends, subsidies, taxes). B_i^t is the amount of bonds paying 1 at time $t+1$, purchased at time t at price $(1+r_t)^{-1}$, where r_t is the interest rate (set by the Central Bank). The maximization is achieved using the standard Lagrange multipliers method over the quantities c_t^i, n_t^i, and B_t^i. This gives the household's state equations Eqs. **3** and **4** and the Euler equation Eq. **5**,[‡]

$$\left(c_t^i\right)^{\varsigma} = \frac{f_t^i}{\lambda_t^i p_t} \qquad [3]$$

$$\left(n_t^i\right)^{\phi} = u_t p_t \frac{\lambda_t^i}{\gamma_i} \qquad [4]$$

$$f_t^i \left(c_t^i\right)^{-\varsigma} = \beta(1+r_t)\mathbb{E}_t\left[\frac{f_{t+1}^i \left(c_{t+1}^i\right)^{-\varsigma}}{1+\pi_{t+1}}\right], \qquad [5]$$

where $u_t = w_t/p_t$ is the wage expressed in price units, $\pi_t := p_t/p_{t-1} - 1$ the inflation rate, and λ_t^i a Lagrange multiplier. The total consumption is $\sum_{i=1}^M c_t^i := C_t$ and the total number of work hours is $\sum_{i=1}^M n_t^i := N_t$.

The unique firm has a technology such that its production Y_t is given by

$$Y_t = M^{\alpha} z_t \frac{N_t^{1-\alpha}}{1-\alpha}, \qquad [6]$$

where z_t is the total factor productivity and α another parameter, often chosen to be $1/3$. The scaling factor M^α is there to ensure a correct limit when $M \to \infty$ that allows total production and consumption must be both proportional to the number of households M. We write $z_t := \bar{z} e^{\xi_t}$, where the log-productivity ξ_t is assumed to follow an autoregressive AR(1) process

$$\xi_t = \eta \xi_{t-1} + \sqrt{1-\eta^2}\,\mathcal{N}\left(0, \sigma^2\right), \qquad [7]$$

where η modulates the temporal correlations of the technology shocks, and σ is the amplitude of these shocks.

Each time period the firm maximizes its profit with the assumption that markets will clear, i.e., that $Y_t \equiv C_t$. Profit is given by $\mathbb{P}_t := p_t C_t - w_t N_t$. Maximization of $\mathbb{P}_t$ yields $u_t = z_t(M/N_t)^{\alpha}$; i.e., the firm hires labor up to the point where its marginal profit equals the real wage (1). Now, assuming for simplicity that f_t^i and γ_i are all equal (homogeneous beliefs and preferences) leads to $c_t^i = c_t = C_t/M$, $n_t^i = n_t = N_t/M$, $\gamma_i = \gamma$, and $f_t^i = f_t$. We can use Eqs. **3**, **4**, and **6** with $Y_t = C_t$ to find c_t, n_t, and u_t as a function of f_t and z_t. In the following, we choose standard values $\phi = \varsigma = 1$, $\alpha = 1/3$, yielding[§]

$$c_t = z_t \left(\frac{9 f_t}{4\gamma}\right)^{1/3}. \qquad [8]$$

Animal Spirits and Self-Reflexivity

Now, the main innovation of the present work is to assume that the sentiment of households at time t (which impacts their consumption propensity f_t) is a function of the past realized consumption of others, itself revealed by a consumer sentiment index. If household i observes that other households have reduced their consumption in the previous time step, it interprets it as a sign that the economy may be degrading (15–17). This increases its precautionary savings and reduces its consumption propensity. Conversely, when other households have increased their consumption, confidence of household i increases, together with its consumption propensity.[¶] A general specification for this animal spirits feedback is

$$f_t^i \longrightarrow F\left(\sum_{j=1, j\neq i}^{M} J_{ij}\, c_{t-1}^j\right), \qquad [9]$$

where $F(.)$ is a monotonic, increasing function and J_{ij} weighs the influence of the past consumption of household j on the confidence level of i. In this work, we consider the case where $J_{ij} = J/M$; i.e., only the consumption "index" matters.[#] This corresponds to a mean-field approximation in statistical physics.

‡Although the Euler equation is actually irrelevant for most of our story, we use it in *Inflation and Narrative-Based Monetary Policy* when we provide an approximate calculation of inflation in the presence of confidence effects.

§Other values can of course be considered as well, but do not change the qualitative conclusions of this paper.

¶In this work, we assign confidence collapse to households, i.e., to the demand side. However, one could argue that confidence collapse in 2008 initially affected the supply side. One can of course easily generalize the present framework to account for such a "wait and see" effect as well. For other attempts to include behavioral effects in DSGE, see ref. 35.

#We furthermore consider the large M limit such that $J/M \sum_{j=1, j\neq i}^{M} c_{t-1}^j \to J c_{t-1}$.

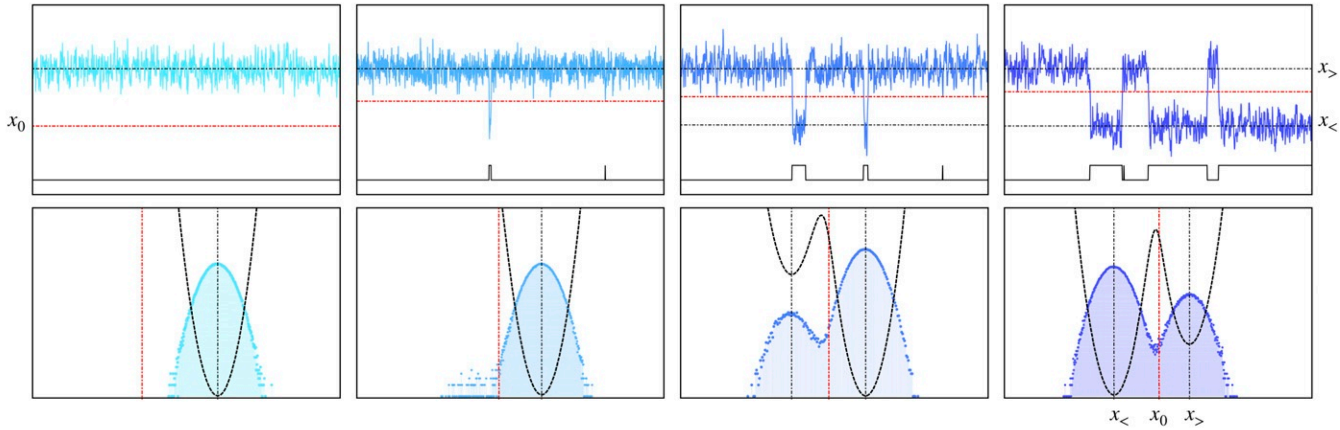

Fig. 1. Numerical simulation of the model for increasing values of the confidence threshold c_0 and for fixed values of $\theta = 5$, $\sigma = 0.6$, and $\eta = 0.5$. (*Top*) Temporal trajectories of the log output $x_t := \log c_t$ with a horizontal dotted-dashed red line located at x_0 and dashed black lines at $x_{>,<}$. (*Bottom*) (Log-)probability distribution $p(x)$ of the output, with the corresponding positions of x_0 and $x_{>,<}$. (*Left* to *Right*) $e^{x_0} = 0.1$ (*A* phase, no crises, Gaussian distribution of output); $e^{x_0} = 0.55$ (B^+ phase, short crises, increased volatility and skewed distribution of output); $e^{x_0} = 0.75$ (*C* phase, long recessions, bimodal distribution with most weight on $x_>$); $e^{x_0} = 1.05$ (*C* phase, long recessions, bimodal distribution with most weight on $\log x_<$). Dashed lines: effective potential $2V(x)/\sigma^2$, defined in *A Theory for Transition Rates*.

While it neglects local network effects, it captures the gist of the mechanism we want to illustrate and furthermore allows us to keep the household homogeneity assumption (different local neighborhoods generally lead to different consumption propensities).

Combining Eqs. **8** and **9** yields

$$c_t = e^{\xi_t} G(c_{t-1}), \quad \text{with} \quad G(x) := \bar{z}\left(\frac{9F(x)}{4\gamma}\right)^{1/3}. \qquad \textbf{[10]}$$

Eq. **10** is a discrete time evolution equation for the consumption level. To exhibit how these dynamics can generate excess volatility and endogenous crises, we assume that $G(x)$ is a shifted logistic function (but any S-shaped function would lead to qualitatively similar results).[‖] To wit, we choose

$$G(c) = c_{\min} + \frac{c_{\max} - c_{\min}}{1 + e^{2\theta(c_0 - c)}}, \qquad \textbf{[11]}$$

where $c_{\min}$, $c_{\max}$, c_0, and θ are parameters with the following interpretation:

- $c_{\min} > 0$ is the minimum level of goods that households will ever consume when productivity is normal (i.e., $\xi_t = 0$).
- $c_{\max} > c_{\min}$ is the maximum level of goods that households will ever consume when productivity is normal (i.e., $\xi_t = 0$).
- c_0 is a "confidence threshold," where the concavity of $G(c)$ changes. Intuitively, $c > c_0$ tends to favor a high-confidence state and $c < c_0$ a low-confidence state.
- $\theta > 0$ sets the width over which the transition from low confidence to high confidence takes place: In the limit $\theta \to +\infty$, one has $G(c < c_0) = c_{\min}$ and $G(c > c_0) = c_{\max}$.[††]

The standard DSGE model, where the animal spirit feedback is absent, is recovered in the limit $\theta c_0 \to -\infty$, in which case $G(c) \equiv c_{\max} = cst$. The dynamics of our extended model fall into four possible phases that we call A, B^+, C, and B^- (Figs. 1 and 2), and we discuss their properties in turn. In the following, we use the notation $\Delta := c_{\max} - c_{\min}$.

Phase Diagram

Phase *A*: High Output, No Crises. This phase corresponds to the DSGE phenomenology, where the only solution of $e^{\xi} G(c) = c$ is a high-consumption solution $c > c_0$ for all values of ξ. Even for large negative shocks $\xi < 0$, the economy remains in a confident state. For small noise amplitude $\sigma \ll 1$, the consumption remains around the value $c_>$ solution of $G(c_>) = c_>$, and one can linearize the dynamics around that point:

$$\delta_{t+1} \approx G'(c_>)\delta_t + \xi_t, \qquad \delta_t := \frac{c_t - c_>}{c_>}. \qquad \textbf{[12]}$$

This leads to the following expression for the consumption volatility:

$$\mathbb{V}[\delta] = \frac{\sigma^2}{1 - G_>'^2}\frac{1 + \eta G'_>}{1 - \eta G'_>}, \qquad G'_> := G'(c_>). \qquad \textbf{[13]}$$

In other words, the output volatility is proportional to the amplitude σ of the technology shocks—small shocks lead to small volatility (note that $G'_> < 1$ in the whole A phase). However, the feedback mechanism leads to excess volatility, since as soon as $G'_> > 0$, one has

$$\frac{\sigma^2}{1 - G_>'^2}\frac{1 + \eta G'_>}{1 - \eta G'_>} > \sigma^2. \qquad \textbf{[14]}$$

The relative position of the boundary of the A phase depends on whether $\theta\Delta$ is larger or smaller than 2. In the first case—corresponding to Fig. 2, the boundary is with the B^+ phase (described below). In the (c_0, θ) plane, the A phase is located on the left of the hyperbola defined by

$$\theta c_0 = 1 + \frac{2 c_{\min}}{\Delta}. \qquad \textbf{[15]}$$

In the case $\theta\Delta < 1$, the boundary is with the B^- phase and the A phase corresponds to $c_0 \le c_{\min} + \Delta/2$.

[‖] Whereas the existence of a feedback is clear (15–17), we are not aware of studies attempting to quantitatively measure the function G. The change of concavity, however, seems plausible from a behavioral viewpoint.

[††] In this work we fix θ and vary c_0. Note, however, that fixing c_0 and varying the "temperature" θ would also be of interest to investigate the effect of population heterogeneity. As mentioned earlier, we leave the study of household heterogeneity for a future work.

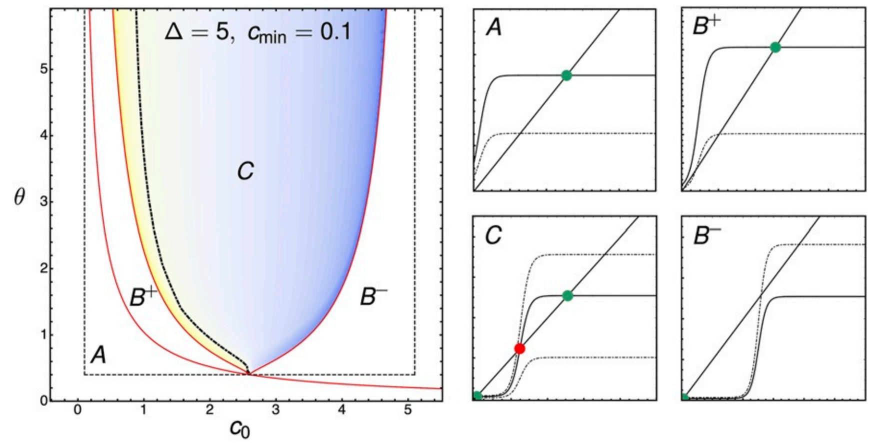

Fig. 2. (*Left*) Phase diagram of the model, with analytically determined boundaries: phase A, high output, no crises; phase B^+, high output with short-lived recessions; phase C, long-lived booms and recessions; phase B^-, low output with short-lived spikes. The color level encodes the distance ratio $(c_> - c^*)/(c^* - c_<)$. This ratio is large in the yellow region, small in the blue region, and equal to one along the dotted-dashed black line. (*Right*) Graphical representation of the iteration $c_{t+1} = e^{\xi_t} G(c_t)$ in the different phases. The solid line corresponds to $\xi_t = 0$, and the dotted–dashed line represents possible realizations with nonzero noise. Clearly, any S-shaped function G would lead to similar effects.

Phase B^+: High Output with Short-Lived Recessions. In this phase B^+ there is still a unique equilibrium state when productivity is normal, i.e., a unique solution to $G(c_>) = c_>$ with $c_> > c_0$. However, downward fluctuations of productivity can be strong enough to give birth to two more solutions $c_< < c^* < c_0$, one unstable (c^*) and one stable ($c_<$). With some exponentially small probability when $\sigma \to 0$ (see Eq. **16** below), the economy can be driven out of the normal state $c_>$ and crash into a low-output state, in which it will remain trapped for a time of the order of $T_\eta := 1/|\ln(\eta)|$, i.e., the autocorrelation time of ξ_t. In other words, sufficiently large fluctuations of output are initially triggered by a relatively mild drop of productivity which is then amplified by the self-referential "panic" effect. But since the low-output state is only a transient fixed point, the recession is only short-lived.

Phase C: Long-Lived Booms and Recessions. Phase C is such that equation $G(c) = c$ has two stable solutions $c_<, c_>$ and one unstable solution c^*. This phase is delimited, in the (c_0, θ) plane, by a parabolic boundary (Fig. 2) with $c_0 \to (c_{\min} + c_{\max})/2$ when $\theta\Delta \to 2^+$ and $c_0 \to c_{\min}$ or $c_{\max}$ when $\theta \to \infty$. The lower boundary $C \to B^+$ corresponds to $c_< \to c^*$ before both disappear, leaving $c_>$ as the only solution, whereas the upper boundary $C \to B^-$ corresponds to $c_> \to c^*$ before both disappear, leaving now $c_<$ as the only solution.

In the absence of fluctuations ($\sigma = 0$), the economy in phase C settles either in a low-output state or in a high-output state. But any, however small, amount of productivity fluctuations is able to induce transitions between these two states. The time needed for such transitions to take place is, however, exponentially long when $\sigma \to 0$,

$$\log T(c_{>,<} \to c_{<,>}) = \frac{W(c_{>,<} \to c_{<,>})}{\sigma^2} + O(\sigma^0), \qquad [16]$$

where $W(c_> \to c_<)$ and $W(c_< \to c_>)$ are computable quantities (*A Theory for Transition Rates* and Fig. 3). This is clearly the most interesting regime: The economy can remain for a very long time in a high-output state $c_>$, with relatively mild fluctuations (in fact still given by Eq. **13**), until a self-fulfilling panic mechanism throws the economy in a crisis state where output is low ($c_<$). This occurs with a Poisson rate $1/T(c_> \to c_<)$. Unless some explicit policy is put in place to restore confidence, the output will linger around $c_<$ for a Poisson time $\sim T(c_< \to c_>)$ which is also very long when $\sigma \to 0$.[‡‡] Note that $T(c_> \to c_<)$ is the average time the system remains around $c_>$ before jumping to $c_<$. The actual time needed to transit is itself short, and the resulting dynamics are made of jumps between plateaus (Fig. 1). A downward jump therefore looks very much like a "crisis."

As we discuss below, recession durations are much shorter than the time between successive crises when $c^* - c_< < c_> - c^*$, i.e., when the low-output solution is close to the unstable solution, which plays the role of an escape point (see below). As c_0 grows larger, one will eventually be in a situation where $c^* - c_< > c_> - c^*$, in which case recession periods are much longer than boom periods (Fig. 2). As σ grows larger, the output flipflops between $c_<$ and $c_>$ at an increasingly faster rate (Fig. 3. While it becomes more and more difficult to distinguish crisis periods from normal periods, the output volatility is dramatically amplified by the confidence feedback loop.

Phase B^-: Low Output with Short-Lived Spikes. Phase B^- is the counterpart of phase B^+ when c_0 is to the right of the phase boundary. In this case, the only solution to $G(c) = c$ is $c_<$: Confidence is most of the time low, with occasional output spikes when productivity fluctuates upward. These output peaks are, however, short-lived and again fixed by the correlation time T_η.[§§].

Phase Diagram: Conclusion. Although quite parsimonious, our model is rich enough to generate a variety of realistic dynamical behaviors, including short-lived downturns and more prolonged recessions (Fig. 1). We tend to believe that the most interesting region of the phase space is in the vicinity of the B^+–C boundary and that the 2008 GFC could correspond to a confidence collapse modeled by a sudden $c_> \to c_<$ transition.[¶¶] The behavior of the economy in the B^- phase, on the other hand, does not seem to correspond to a realistic situation. One of our major results is that the crisis probability is exponentially sensitive to the parameters of the model.

[‡‡] Note in particular that $T(c_< \to c_>)$ is much longer than T_η: It is no longer the correlation time of productivity fluctuations that sets the duration of recessions (at variance with the B^+ scenario).

[§§] Note that there is no "A^-" analogue of the A phase described above—this is due to the fact that the productivity factor z_t can have unbounded upward fluctuations but cannot become negative.

[¶¶] The role of trust in the unraveling of the 2008 crisis is emphasized in Bernanke et al. (36).

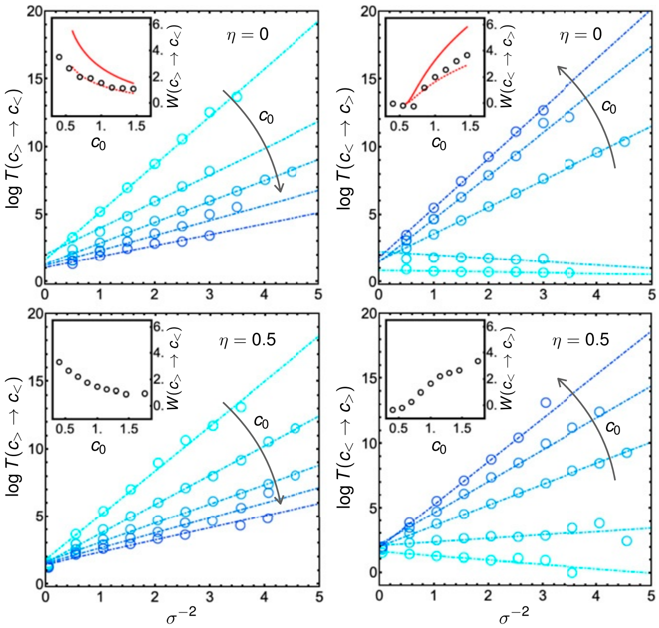

Fig. 3. Plot of $\log T(c_> \to c_<)$ (*Left* column) and $\log T(c_< \to c_>)$ (*Right* column) vs. σ^{-2} for different values c_0 and $\eta = 0$ (*Top* row) and $\eta = 0.5$ (*Bottom* row). The value of c_0 increases with the circles' tonality becoming darker. The linear dependence confirms the validity of ref. 16. *Insets* show the corresponding barriers W as a function of c_0. For $\eta = 0$, we plot the continuous-time prediction (20) with $\varepsilon = 1$ (solid red line), which overestimates the true barriers (dotted red line) by a factor ≈ 2.

A Theory for Transition Rates

Discrete Maps. Let us now discuss in more detail one of the most important predictions of our model, namely the exponential sensitivity to σ of the crisis probability, Eq. **16**. Such a result can be obtained by adapting the formalism of ref. 37 to the present problem. In terms of $x_t := \log c_t$, the map (10) reads

$$x_t = H(x_{t-1}) + \xi_t, \qquad \textbf{[17]}$$

with $H(x) := \log G(e^x)$. In the limit of white noise (i.e., $\eta = 0$ in Eq. **7**), this is precisely the general problem studied in ref. 37 in the case where $H(x) = x$ has two stable solutions and an unstable one in between. The authors of ref. 36 show that the average time before jumping from one stable solution to another is given, for small σ, by Eq. **16**. They provide an explicit scheme to compute (at least numerically) the quantity W, called the activation barrier in physics and chemistry. The idea is to find the most probable configuration of ξ_ts that allows the system to move from one stable position to another. In a nutshell, this amounts to finding a heteroclinic connection, in an enlarged space, between the starting point and the intermediate, unstable fixed point $x^* = \log c^*$ (38).

It is straightforward to generalize the approach of ref. 37 and see that the jump rate has the same exponential dependence on σ^2 when the correlation time T_η is nonzero, as confirmed by Fig. 3. However, finding the value of W is more complicated. Approximation methods can be devised in the continuous-time limit that we describe now.

Continuous-Time Limit. Let us slightly change the dynamics by assuming that x_t depends not on the previous value x_{t-1} but rather on an exponential moving average $\bar{x}_{t-1}$ of past values of x, defined recursively as

$$\bar{x}_{t-1} = (1-\varepsilon)\bar{x}_{t-2} + \varepsilon x_{t-1}. \qquad \textbf{[18]}$$

Eq. **17** instead reads $x_t = H(\bar{x}_{t-1}) + \xi_t$. Eliminating x_t yields

$$\bar{x}_t - \bar{x}_{t-1} = \varepsilon\left(H(\bar{x}_{t-1}) - \bar{x}_{t-1} + \xi_t\right). \qquad \textbf{[19]}$$

In the limit $\varepsilon \to 0$, Eq. **19** becomes a Langevin equation (or stochastic differential equation) for $\bar{x}_t$, for which a host of results are available. It is useful to introduce a potential function $V(x)$ such that $V'(x) = x - H(x)$. The potential $V(x)$ has two minima ("valleys") corresponding to $\log c_<$ and $\log c_>$ and a maximum ("hill") corresponding to $\log c^*$ (Fig. 1, *Bottom* row). With this representation, the dynamics of $\bar{x}_t$ under Eq. **19** become transparent: For long stretches of time, $\bar{x}_t$ fluctuates around either $x_< = \log c_<$ or $x_> = \log c_>$, until rare fluctuations of ξ_t allow the system to cross the barrier between the two valleys. Calculating the rate Γ of these rare events is the classic Kramers problem (for a comprehensive review, see ref. 39). In the limit $T_\eta = 0$ where the noise is white, the final exact expression is, for $\sigma \to 0$,

$$\Gamma(x_> \to x_<) = \frac{\sqrt{|H'(x_>)H'(x^*)|}}{2\pi} \exp\left(-\frac{2W}{\varepsilon\sigma^2}\right),$$
$$W := V(x^*) - V(x_>), \qquad \textbf{[20]}$$

and mutatis mutandis for $\Gamma(x_< \to x_>)$. Such a prediction is compared with numerical simulations in Fig. 3; it overestimates the real barrier by a factor ≈ 2. The most important feature is the exponential dependence of this rate on the height of the barrier W and on the inverse noise variance σ^2.[##]

Exponential Dependence and "Unknown Knowns." It is worth emphasizing the economic consequences of this exponential dependence of the probability of crises in our model. Clearly, any small uncertainty about the parameters of the model (i.e., $c_0, c_{\min}, c_{\max}, \theta$) or for that matter the precise specification of the function $G(c)$, or any other feature neglected in the model, will no doubt affect the precise value of the barrier W. But in the rare-event regime $W/\sigma^2 \gg 1$ any uncertainty on W is exponentially amplified. Take for example $W/\sigma^2 = 25$; a small relative error of 10% on W changes the crisis rate by one order of magnitude. Precisely as the famous butterfly effect (i.e., the exponential sensitivity on initial conditions) forbids any deterministic description of chaotic systems, the exponential dependence of the crisis rate means that this rate is, for all practical purposes, unknowable. Since the probability of rare events cannot be determined empirically, it means that no market can provide a rational valuation of the corresponding risks. This is an interesting example of "unknown knowns," where what may happen is known, but its probability is impossible to quantify and cannot be priced.

Inflation and Narrative-Based Monetary Policy

In the absence of frictions, the model is usually closed by assuming a Taylor rule for the interest rate, as $r_t = \Phi\pi_t - \log\beta$, with $\Phi > 1$ fixing the amplitude of the response of the Central Bank to inflation π (1). Let us first assume that the crisis probability is very small, so one can linearize the Euler equation Eq. **5** with $\varsigma = 1$. Solving forward in time leads to

$$\pi_t + \frac{\kappa_>}{\Phi}(\delta_t - \delta_{t-1}) = \left(1 - \frac{\kappa_>}{\Phi}\right)\sum_{k=0}^{\infty}\Phi^{-k-1}\mathbb{E}_t[\delta_{t+k+1} - \delta_{t+k}], \qquad \textbf{[21]}$$

where δ_t is the output gap defined in ref. 12 and $\kappa_> := 3G'(c_>) \geq 0$. In the limit $\kappa_> \to 0$ and one recovers the standard expression (1). The self-reflexive feedback adds a term

[##] The generalization of Kramers' result for so-called colored noise (i.e., $T_\eta > 0$) is also available (40). In this case, corrections to Eq. **20** can be systematically computed, but the exponential dependence of Γ on σ^{-2} is preserved.

that depends on the past output gap trend and changes the coefficient in front of the expected future output gap variations. Interestingly, $1-\kappa_>/\Phi$ can become negative for some range of parameters. Accounting for crises analytically is difficult in general, since the Euler equation cannot be linearized anymore. To make progress, we model the dynamics as follows: With probability $p=T^{-1}(c_> \to c_<) \ll 1$ the economy crashes between t and $t+1$, and with probability $1-p$ it hovers normally around $c_>$, with small fluctuations. We also assume that $\pi_t \ll 1$. Hence, we approximate the right-hand side of the Euler equation Eq. **5** as

$$\frac{F(c_t)}{c_>}(1+\Phi\pi_t)\left((1-p)\mathbb{E}_t^>[(1-\pi_{t+1}-\delta_{t+1})]+p\frac{c_>}{c_<}\right), \quad [22]$$

where $\mathbb{E}_t^>$ is an expectation conditional to remaining near the high-output equilibrium. This eventually leads to an extra term in Eq. **21** equal to

$$\delta\pi_t = -\frac{p}{\Phi-1}\,\frac{c_> - c_<}{c_<}. \quad [23]$$

As expected, anticipation of possible crises decreases inflation; provided $c_< \ll c_>$ this correction can be substantial even when $p \ll 1$.

Our setting corresponds, up to now, to a proto-DSGE model. Including frictions (like Calvo's staggered price adjustment) would lead to a richer model, with, for example, a modified "new Keynesian Phillips curve" (1). Of particular importance would be to include market breakdown in crisis periods, i.e., allowing for $C_t \neq Y_t$: Production and consumption will not match as confidence collapses. We leave these extensions for future research.

In our opinion, however, the most important aspect of our model is that it justifies alternative, behavioral tools for monetary policy, in particular in crisis time. Beyond adjusting interest rates and money supply, policymakers could also use narratives to restore trust,[†††] parameterized in our model by the threshold c_0. If the economy lies in the neighborhood of the C/B^+ phase boundary (Fig. 2), a mild decrease of c_0, engineered by the Central Bank, may help in putting the system back on an even keel.

The authors declare that they have no conflict of interest. There are no empirical data in this paper, just analytical and numerical solving of some equations that the reader can easily compute to reproduce the results.

ACKNOWLEDGMENTS. We thank A. Amstrong, R. Farmer, X. Gabaix, S. Gualdi, A. Kirman, J. Scheinkman, and F. Zamponi for many insightful discussions on these topics. This research was conducted within the Econophysics & Complex Systems Research Chair, under the aegis of the Fondation du Risque, the Fondation de l'Ecole Polytechnique, the Ecole Polytechnique, and Capital Fund Management and is part of the National Institute of Economic and Social Research (UK) Rebuilding Macroeconomics initiative. M.T. is a member of the Institut Universitaire de France.

[†††]The importance of narratives in economics was recently stressed in ref. 41.

1. J. Galí, *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework and Its Applications* (Princeton University Press, 2015).
2. W. H. Buiter, The unfortunate uselessness of most 'state of the art' academic monetary economics. *VoxEU* (2009). https://voxeu.org/article/macroeconomics-crisis-irrelevance. Accessed 4 April 2020.
3. Trichet, J.-C., Reflections on the nature of monetary policy non-standard measures and finance theory: Speech by Jean-Claude Trichet, President of the ECB, Opening address at the ECB Central Banking Conference, Frankfurt, 18 November 2010. https://www.ecb.europa.eu/press/key/date/2010/html/sp101118.en.html. Accessed 4 April 2020.
4. O. Blanchard, Where danger lurks. *Finance Dev.* **51**, 28–31 (2014).
5. J. E. Stiglitz, Where modern macroeconomics went wrong. *Oxf. Rev. Econ. Pol.* **34**, 70–106 (2018).
6. L. J. Christiano, M. S. Eichenbaum, M. Trabandt, On DSGE models. *J. Econ. Perspect.* **32**, 113–140 (2018).
7. O. Blanchard, On the future of macroeconomic models. *Oxf. Rev. Econ. Pol.* **34**, 43–54 (2018).
8. R. Reis, Is something really wrong with macroeconomics?*Oxf. Rev. Econ. Pol.* **34**, 132–155 (2018).
9. D. Gatti, E. Gaffeo, M. Gallegati, G. Giulioni, A. Palestrini, *Emergent Macroeconomics: An Agent-Based Approach to Business Fluctuations* (Springer Science & Business Media, 2008).
10. G. Fagiolo, A. Roventini, Macroeconomic policy in DSGE and agent-based models. *Rev. OFCE* **124**, 67–116 (2012).
11. S. Gualdi, M. Tarzia, F. Zamponi, J.-P. Bouchaud, Tipping points in macroeconomic agent-based models. *J. Econ. Dynam. Contr.* **50**, 29–61 (2015).
12. C. Hommes, B. LeBaron, *Computational Economics: Heterogeneous Agent Modeling* (Elsevier, 2018).
13. G. Dosi, A. Roventini, More is different…and complex! the case for agent-based macroeconomics. *J. Evol. Econ.* **29**, 1–37 (2019).
14. A. G. Haldane, A. E. Turrell, An interdisciplinary model for macroeconomics. *Oxf. Rev. Econ. Pol.* **34**, 219–251 (2018).
15. S. Celik, Y. Ozerkek, Panel cointegration analysis of consumer confidence and personal consumption in the European Union. *J. Bus. Econ. Manag.* **10**, 161–168 (2009).
16. C. Christiansen, J. N. Eriksen, S. V. Moller, Forecasting US recessions: The role of sentiment. *J. Bank. Finance* **49**, 459–468 (2014).
17. A. M. Klopocka, Does consumer confidence forecast household saving and borrowing behavior? Evidence for Poland. *Soc. Indicat. Res.* **133**, 693–717 (2017).
18. B. Bernanke, M. Gertler, S. Gilchrist, The financial accelerator in a quantitative business cycle framework. *Handb. Macroeconomics* **1**, 1341–1393 (1999).
19. W. A. Brock, S. N. Durlauf, Discrete choice with social interactions. *Rev. Econ. Stud.* **68**, 235–260 (2001).
20. J. Campbell, J. Cochrane, By force of habit: A consumption-based explanation of aggregate stock market behavior. *J. Polit. Econ.* **107**, 205–251 (1999).
21. J. Gali, Keeping up with the Joneses: Consumption externalities, portfolio choice, and asset prices. *J. Money Credit Bank.* **26**, 1–8 (1994).
22. L. Stracca, A. Al-Nowaihi, Keeping up with the Joneses, reference dependence, and equilibrium indeterminacy. https://ssrn.com/abstract=657863. Accessed 24 March 2020.
23. K. Anand, A. Kirman, M. Marsili, Epidemics of rules, rational negligence and market crashes. *Eur. J. Finance* **19**, 438–447 (2013).
24. João. d. G. Batista, J.-P. Bouchaud, D. Challet, Sudden trust collapse in networked societies. *Eur. Phys. J. B* **88**, 55 (2015).
25. Y. M. Asano, J. J. Kolb, J. Heitzig, J. D. Farmer, Emergent inequality and endogenous dynamics in a simple behavioral macroeconomic model. arXiv:1907.02155 (3 July 2019).
26. J.-P. Bouchaud, S. Gualdi, M. Tarzia, F. Zamponi, Optimal inflation target: Insights from an agent-based model. *Economics*, 10.5018/economics-ejournal.ja.2018-15 (2017).
27. W. Carlin, D. Soskice, Stagnant productivity and low unemployment: Stuck in a Keynesian equilibrium. *Oxf. Rev. Econ. Pol.* **34**, 169–194 (2018).
28. P. D. Grauwe, Animal spirits and monetary policy. *Econ. Theor.* **47**, 423–457 (2011).
29. C. Chiarella, The dynamics of speculative behaviour. *Ann. Oper. Res.* **37**, 101–123 (1992).
30. T. Lux, M. Marchesi, Scaling and criticality in a stochastic multi-agent model of a financial market. *Nature* **397**, 498–500 (1999).
31. M. Wyart, J.-P. Bouchaud, Self-referential behaviour, overreaction and conventions in financial markets. *J. Econ. Behav. Organ.* **63**, 1–24 (2007).
32. A. Majewski, S. Ciliberti, J.-P. Bouchaud, Co-existence of trend and value in financial markets: Estimating an extended Chiarella model. *J. Econ. Dyn. Control* **112**, 103791 (2020).
33. J.-P. Bouchaud, Crises and collective socio-economic phenomena: Simple models and challenges. *J. Stat. Phys.* **151**, 567–606 (2013).
34. R. E. A. Farmer, The indeterminacy agenda in macroeconomics. https://www.nber.org/papers/w25879 (May 2019). Accessed 4 April 2020.
35. X. Gabaix, A behavioral new Keynesian model. https://www.nber.org/papers/w22954 (December 2016). Accessed 4 April 2020.
36. B. Bernanke, T. F. Geithner, H. M. Paulson, *Firefighting: The Financial Crisis and Its Lessons* (Penguin Books, 2019).
37. J. Demaeyer, P. Gaspard, A trace formula for activated escape in noisy maps. *J. Stat. Mech. Theor. Exp.* **2013**, P10026 (2013).
38. J. Gruckenheimer, P. H. Holmes, *Nonlinear Oscillations, Dynamical System, Bifurcation of Vector Fields* (Springer-Verlag, New York, NY,1983).
39. P. Hänggi, P. Talkner, M. Borkovec, Reaction-rate theory: Fifty years after Kramers. *Rev. Mod. Phys.* **62**, 251–341 (1990).
40. A. J. Bray, A. J. McKane, Instanton calculation of the escape rate for activation over a potential barrier driven by colored noise. *Phys. Rev. Lett.* **62**, 493–496 (1989).
41. R. J. Shiller, Narrative economics. *Am. Econ. Rev.* **107**, 967–1004 (2017).