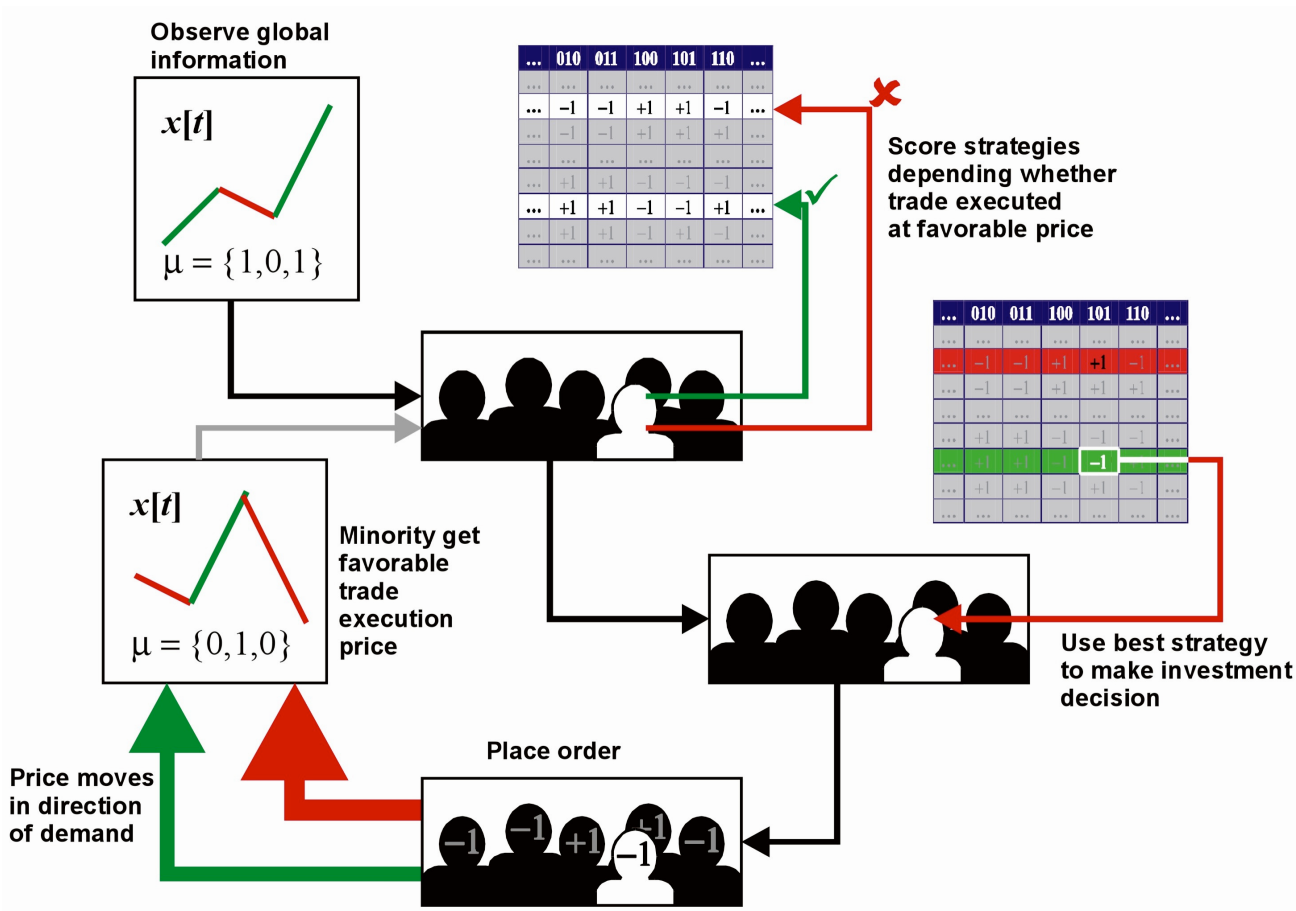
Supplementary Information (SI) for manuscript

"Abrupt rise of new machine ecology beyond human response time"

Author list: Neil Johnson, Guannan Zhao, Eric Hunsader, Hong Qi, Nicholas Johnson, Jing Meng and Brian Tivnan

Contents:

1. Our model used in the main paper: Our model, which is summarized in SI Figure 1, considers a simple yet archetypal model of a complex system such as a financial market, based on a population of agents competing for a limited resource with bounded rationality. This model reproduces well-known stylized facts of financial markets (see Ref. [31] of main paper).



SI Figure 1. Our dynamical market model is based on the realistic notion that it is better to be a buyer when there is an excess of sellers or vice versa when in a financial market comprising agents (humans or machines) with short-term, high-frequency trading goals. Full details are given in Ref. [31]. The length of the price history bit-string is *m*=3, and the number of strategies per agent is *s*=2. Agents are heterogeneous since they each have their own *s* strategies pulled from the space of available strategies (i.e. the strategy space). As shown in the top-left panel, the information input at the beginning of each timestep is the current price history bit-string μ which lists the signs of the price-changes over the previous *m*=3 timesteps, i.e. up-down-up which is encoded in binary form as $\mu = \{1,0,1\}$ where 1 means up and 0

means down. Agents may be human or electronic, and at each timestep each agent adopts his/her own current best (i.e. highest-scoring) strategy from his/her own set of *s* strategies. The current best strategy for the agent shown in white mandates action -1 (i.e. the strategy recommends to sell) in response to the current price-history information $\mu = \{1,0,1\}$. The strategy space is shown as a table, with the top row being the possible price-history bit-strings for *m*=3, and the entries in each subsequent row comprising actions +1 (i.e. buy) and -1 (sell). Hence each row represents a single strategy, i.e. it gives a well-defined response for each possible μ and hence each possible situation. In this example, each agent holds two strategies (i.e. two rows) chosen randomly from all rows at the beginning of the simulation, and hence the agents are heterogeneous. At a given timestep, there is a given price-history bit-string μ (e.g. $\mu = \{1,0,1\}$ as shown); each agent chooses its best performing strategy and hence follows the action shown in the entry under column μ, which in the example shown is -1 (i.e. sell). All the agents follow this same procedure: They choose their own best strategy and receive its recommendation, i.e. buy (+1) or sell (-1). In the example timestep shown, more agents sell (-1) than buy (+1). The new price-change, given by an excess demand $D = (2-4) < 0$ (i.e. 2 buyers given by +1, and 4 sellers given by -1, hence the net demand is -2) is therefore downward, i.e. it gets added to the price history μ as a 0. Therefore the new *m*=3 price history bitstring for the next timestep is $\mu = \{0,1,0\}$ as shown in the middle-left panel. This process then iterates in the same manner for all timesteps.

SI Tables 1 and 2 provide support for our model's mechanism (SI Figure 1) in which it is better to be a buyer when there is an excess of sellers or vice versa when in a financial market comprising agents (humans or machines) with short-term, high-frequency trading goals. This is equivalent to saying it is better to be in the minority group in such a market. Specifically, an agent could be an automated trading platform with *s* being the number of algorithms that this platform manages. We define a notional wealth W_i of an agent i at time t as follows: $W_i[t] = \phi_i[t]x[t] + C_i[t]$ where ϕ_i is the number of assets that it holds at time t, C_i is the amount of cash it holds at time t, and $x[t]$ is the asset price at time t. Since an exchange of cash for assets does not affect the agents' overall wealth at that moment, $W_i[t]$ is a notional wealth. The real measure of wealth is C_i, which is the amount of capital that the agent has available to spend. An agent has to do a 'round trip' (i.e. buy (sell) an asset then sell (buy) it back) to discover whether a real profit has been made.

t	Action $a[t]$	$C_i[t]$	$\phi_i[t]$	$x[t]$	$W_i[t]$
1	submit a buy order	100	0	10	100
2	buy…, submit a sell order	91	1	9	100
3	Sell	**101**	0	10	101

SI Table 1. Winning by trading in the minority group. For the buy action, the agent (human or machine) is a buyer when the majority are sellers. Because of the negative market impact of an excess of sellers, the price $x[t]$ at which the trade is finally executed is below the advertised price $x[t] = 10$ (i.e. it is executed at $x[t] = 9$ in our example and hence only costs the trader 9 units of cash). For the sell action, the trader is a seller when the majority are buyers, hence the price $x[t]$ at which the trade is finally executed is above the advertised price $x[t] = 9$ (i.e. it is executed at $x[t] = 10$ in our example). Hence the agent ends up with 101 units of cash, having started with only 100. The agent has therefore made a profit of 1 unit of cash after the round-trip, simply by trading in the minority group, hence providing support for our model's mechanism (SI Figure 1) where agents try to trade in the minority.

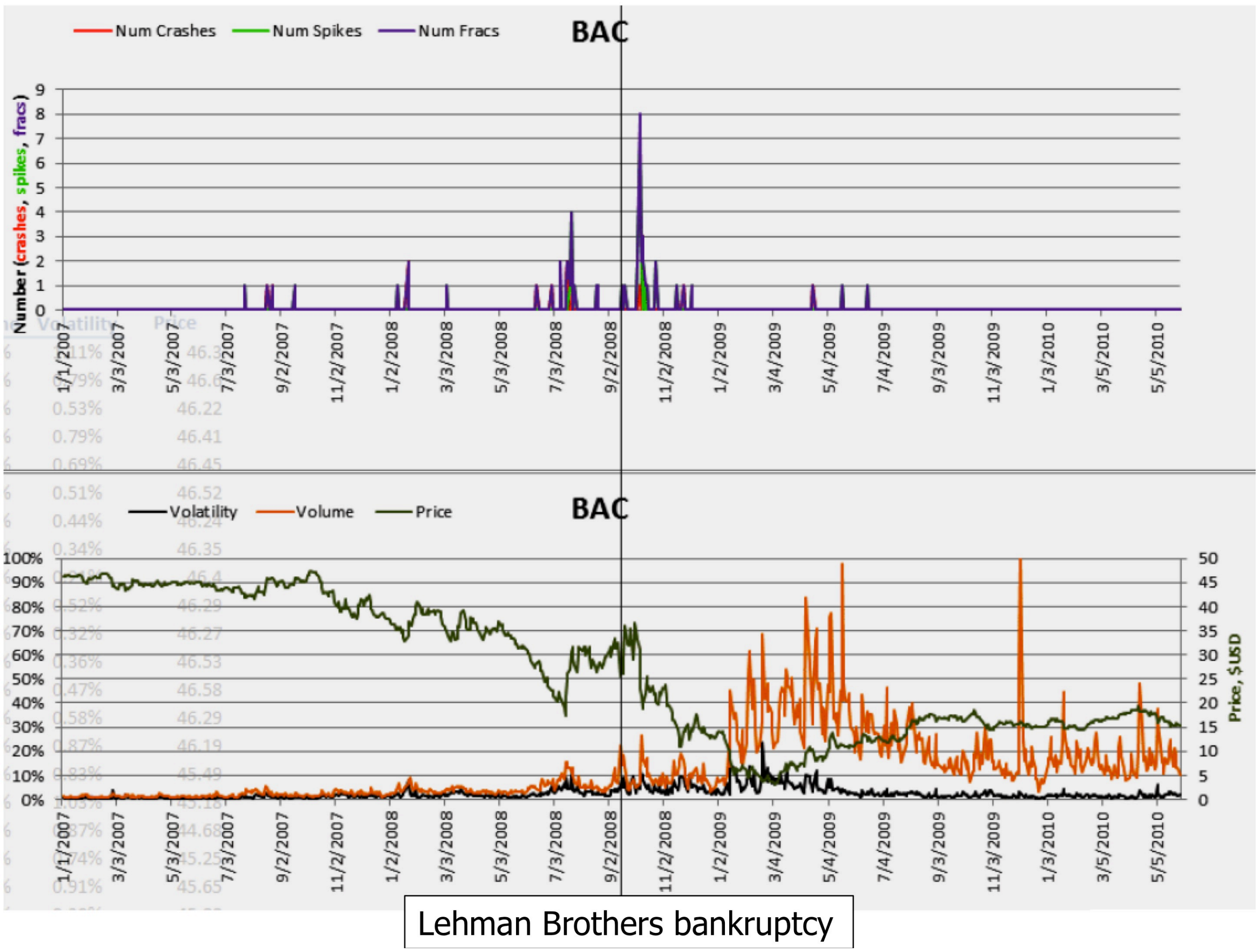
t	Action $a[t]$	$C_i[t]$	$\phi_i[t]$	$x[t]$	$W_i[t]$
1	Submit a buy order	100	0	10	100
2	buy…, submit a sell order	89	1	11	100
3	Sell	**99**	0	10	99

SI Table 2. Losing by trading in the majority group. For the buy action, the agent is a buyer when the majority are also buyers. Because of the positive market impact of an excess of buyers, the price $x[t]$ at which the trade is finally

executed is above the advertised price $x[t]=10$ (i.e. it is executed at $x[t]=11$ in our example and hence costs the trader 11 units of cash). For the sell action, the trader is a seller when the majority are also sellers, hence the price $x[t]$ at which the trade is finally executed is below the advertised price $x[t]=11$ (i.e. it is executed at $x[t]=10$ in our example). Hence the agent ends up with 99 units of cash, having started with 100. The agent has therefore lost 1 unit of cash after the round-trip, simply by trading in the majority group. Being in the majority, as opposed to the minority, is therefore undesirable in such a high-frequency trading scenario – hence our model is reasonable.

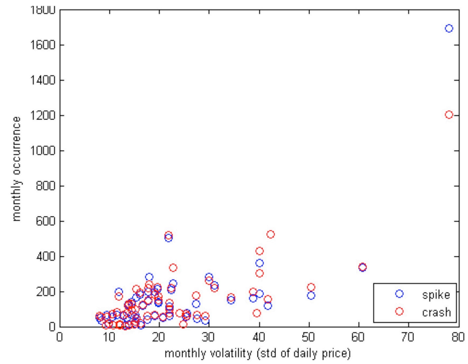
2. Additional supporting results for the claims made in the main paper concerning the nontrivial nature of the ultrafast extreme events (UEEs) and support for a transition around 1 second

The following figure demonstrates that the ultrafast extreme events (UEEs) that we study, are not just a simple consequence or knock-on effect of large daily volume, or volatility, or price movements.

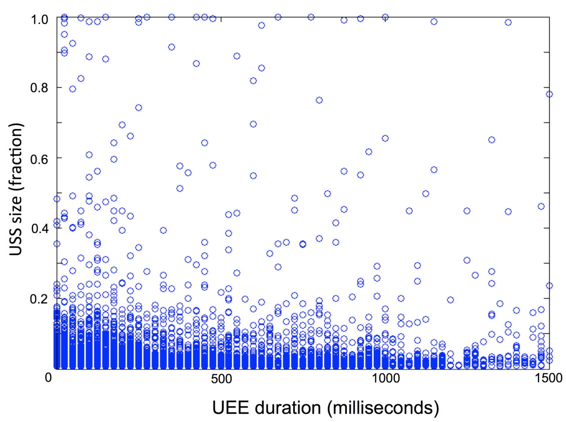


SI Figure 2: Example of (bottom panel) price, volume, volatility, and (top panel) ultrafast spikes and crashes, in Bank of America (BAC) stock around the 2008 financial instability shown in Fig. 1C of the main paper. Top plot suggests an escalation of ultrafast extreme events in the lead-up to the 2008 financial collapse. The start and end times of the escalation period shown in Fig. 1C of the main paper is determined by examining the local trend in the arrival rate of the ultrafast extreme events – we have checked various methods such as LOWESS and found them all to give very similar escalation periods as shown in Fig. 1C. Hence our statements and conclusions are robust. The Lehman Brothers bankrupty event is also indicated by a vertical black line. The total number of ultrafast extreme events (i.e. spikes and crashes combined) is shown as 'Num Fracs' (purple). There is no clear relationship between the appearance of the ultrafast extreme events, and high/low daily volume, volatility or daily price behavior. This confirms our claim that the ultrafast extreme events that we uncover and study in the main paper, are not just a simple consequence or 'knock-on effect' of large daily volume, or volatility, or price movements. Instead, there is a nontrivial coupling across timescales,

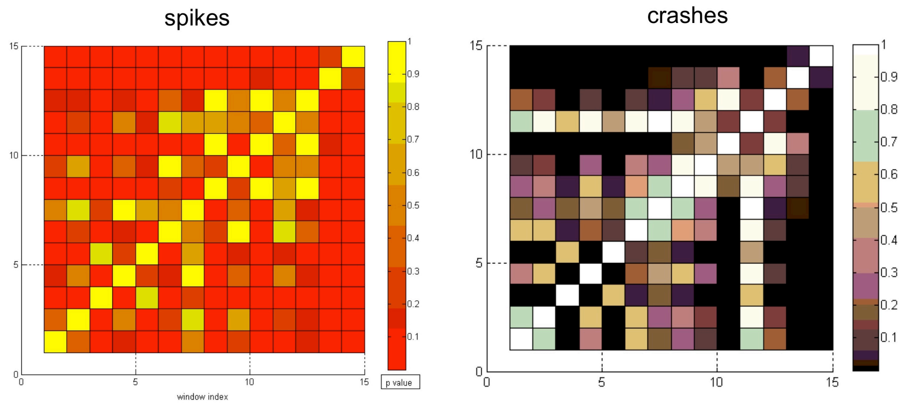
from subsecond to monthly. A similar story holds for other stock charts (not shown). We are grateful to Joel Malerba for producing this plot.



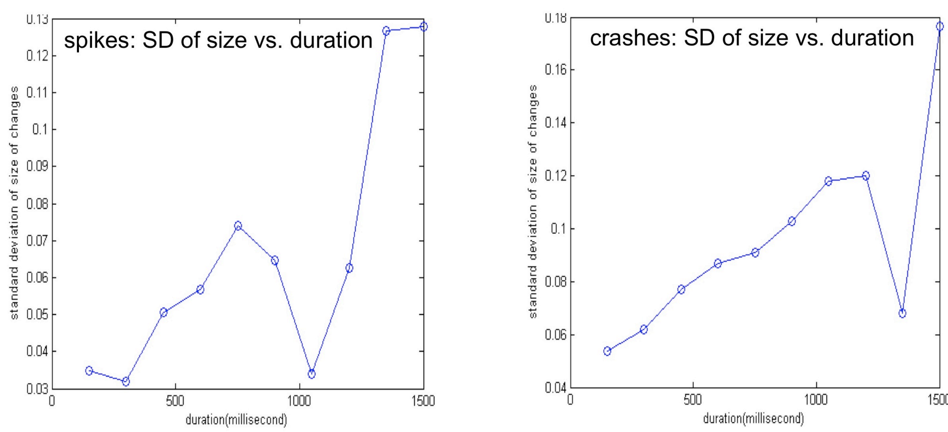
SI Figure 3: Number of ultrafast spikes (blue) and ultrafast crashes (red) in individual months, as a function of the volatility in that month. As can be seen, there is a non-trivial relationship between the number of ultrafast extreme events (i.e. spikes and crashes) and the overall volatility of the market on much longer timescales (i.e. months). This supports our claim that there is an interesting non-trivial coupling across timescales.



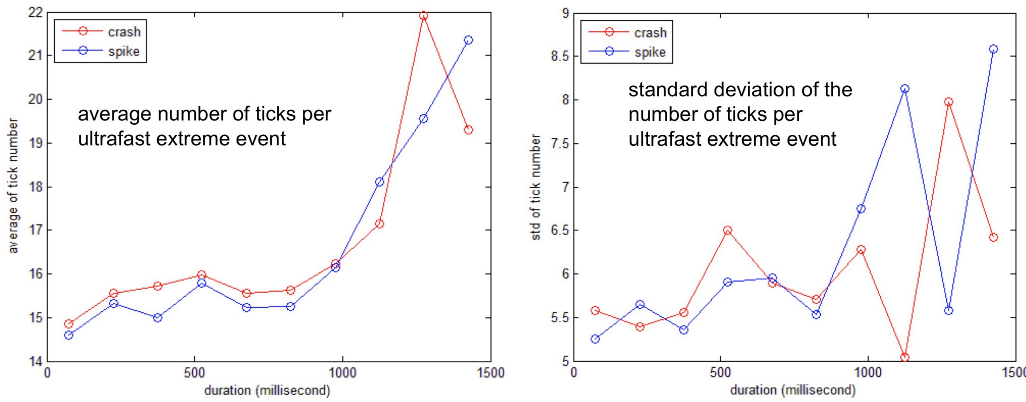
SI Figure 4: Size vs. duration for all ultrafast extreme events including both crashes and spikes. For the sake of clarity in the plot, we truncate events with size larger than 100%, i.e. we have truncated the plot at 100 on the vertical scale. As can be seen, there is no well-defined relationship between spike/crash size and duration. The fact that size and duration of the ultrafast extreme events are not trivially linked helps confirm the surprising nature of these ultrafast extreme events.



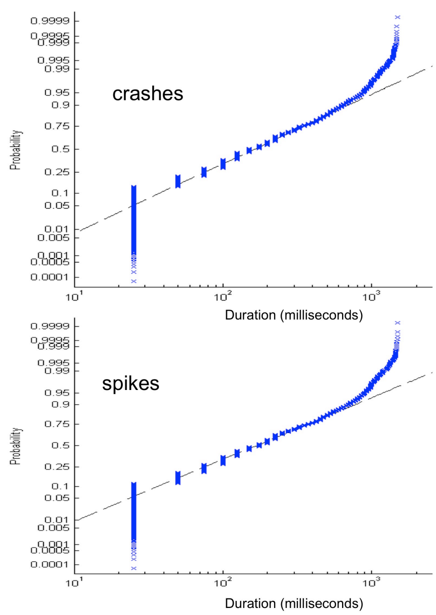
SI Figure 5: Distribution of ultrafast extreme event sizes for fixed, consecutive, non-overlapping time-windows for the duration (e.g. durations between 400-500ms, 500-600ms etc.). The colors (with values indicated) show the results of the Kolmogorov-Smirnov two-sample test to check the similarity of the different distributions within 15 different time-windows. The fact that there is little similarity between distributions on the longer timescale (> 1 second) and ones a few hundred milliseconds below, is consistent with the claim that the ultrafast extreme events of duration below about 800-900ms are fundamentally different from those above, i.e. there is a transition. Crashes show more similarity across timescales than spikes, as expected from their broader transition in Fig. 3 of the main paper. We also carried out the power law test on these same consecutive, non-overlapping time-windows for the duration (e.g. power-law test on durations between 400-500ms, 500-600ms etc.). Only the time-windows containing long duration ultrafast extreme events pass the *p*-value test, which is consistent with our claim of a transition just below the human response time.



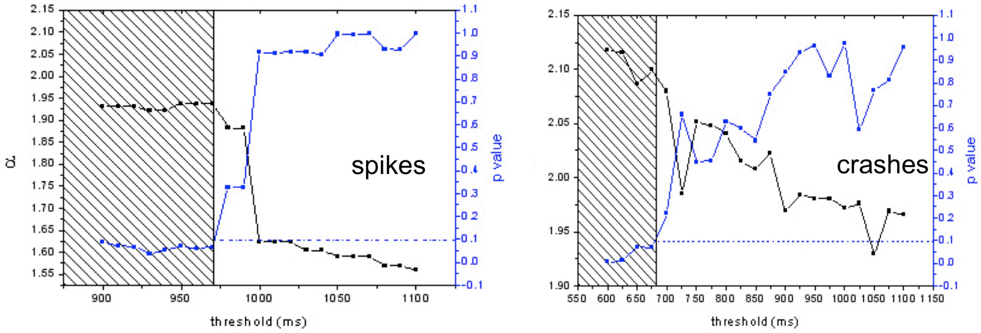
SI Figure 6: Panels show the standard deviation of the size of the ultrafast extreme events within a given window of duration as a function of the upper value of the duration for this window. The windows of duration are fixed, consecutive, non-overlapping, with each one having a size of approximately 150 ms (there are 10 windows in total). Again, there seems to be a transition with the subsecond regime appearing fundamentally different from the regime of >1 second. The corresponding diagram for the mean size (not shown) shows the mean size increasing as the duration moves above 1 second, while the mean number decreases as shown above. Although there are fewer ultrafast extreme events at larger durations, the mean size of them is larger simply because they have more time to develop and grow. The fact the size increases so abruptly is another indication that the underlying distribution has changed. The emergence of a power-law-like distribution above 1 second will generate a large mean for durations >1 second, as expected for such a fat-tailed distribution -- in stark contrast, non-power-law distributions below 1 second will generate a smaller average, exactly as observed.



SI Figure 7: Average and standard deviation in the number of price ticks making up the individual ultrafast extreme events which lie within a given duration window. The rapid changes around 1000ms in each case again support the claim that there is a fundamental difference between the ultrafast extreme events in the subsecond and >1 second regimes. The window (bin) size for durations is 150ms.



SI Figure 8: Lognormal distribution (straight dashed line) fit to complementary distribution function for ultrafast extreme event durations. The large deviation from the straight line around 1000 milliseconds, where the straight line represents a lognormal distribution, provides evidence supporting our claim of a transition point for durations around 1000 milliseconds (i.e. 1 second).



SI Figure 9: Plot shows that the values and curves in Fig. 3 of the main paper, retain the same form – and hence the transition that we identify is robust – even if the durations are binned in a different way. Figure 3 is effectively a smoothed version of this figure.

3. How to access more detailed information about each ultrafast extreme event (UEE) through the NANEX website

In order to access the individual data for each individual UEE, we provide the following guide. The data are consistent with the claim that the UEEs do not result from one single trade, nor are they instantaneous, though they of course appear rather abrupt on the horizontal scale of seconds used in the plots. There are in general trades of irregular sizes, occurring at irregular prices, throughout the duration of the UEEs. Instead, the data is consistent with a sudden crowd of trades, which in turn is consistent with the mechanism of our model whereby a crowd of agents (computers) starts to emerge using the same strategy.

Steps to access data for each UEE:

1. Go to http://www.nanex.net
2. Click on Nanex Research to access page http://www.nanex.net/FlashCrash/OngoingResearch.html
3. In column "Research" on left hand side, click "Micro Flash Crashes" to access http://www.nanex.net/FlashCrashEquities/FlashCrashAnalysis_Equities.html
4. This contains all the UEEs from 2006 onwards. These were the dataset used in the study.
5. As stated on this page, the links on this page contain "…ZIP archives for each year analyzed. Simply download the files, unzip and start viewing."
6. As examples, there are also 10 pages with 10 sample images from each year to view.
7. As stated, each chart contains a set of numbers in the upper left corner, which give the details for each.

Quoting from this NANEX-produced webpage at www.nanex.net (courtesy of NANEX):
"The charts presented in this report show trading activity in 250ms increments. Time is is shown on the X-axis both on the top and bottom of the chart. Price is shown on the Y-axis on the left and right of the chart. Each trade is represented by a colored dot, the color designating the exchange the trade was reported from. The legend for the exchanges by number is:
1) Nasdaq
2) Nasdaq Alternative Display Facility
3) New York Stock Exchange
4) American Stock Exchange
5) Chicago Board Options Exchange
6) International Securities Exchange
7) NYSE Arca
8) National Stock Exchange (Cincinnati)

9) Philidelphia Exchange
11) Boston
17) Chicago Stock Exchange
57) Nasdaq Trade Reporting Facility
59) NYSE Trade Reporting Facility
60) BATS Trading
64) Direct Edge
65) Direct Edge X
The dots in the actual price chart are sized according to the size of the trade.
The dots in the exchange legend are sized according to the frequency of their appearance on the chart.
Within each chart in the upper left is a black box with white text. The lines of text represent:
1) Exg: Symbol
2) Date
3) Number of sequential up or down ticks and price change.
4) Duration and price change percentage."
This information is consistent with the claim that that the UEEs are not instantaneous, nor are they coming from one trade. We thank NANEX for their courtesy in providing us access to this data.