Supporting Information

Moussaïd et al. 10.1073/pnas.1421883112

SI Materials and Methods

Coding of the Message Content. The distortion of the message content describes how accurately the "factual" units of information were transmitted from one subject to the next and to what extent these information units change as result of social transmission. The distortion of the content is mostly neutral with respect to the signal of the message; these changes are due to, for instance, inaccurate recall or careless and unintentional wording. For example, when the information "Up to 0.3 mg of triclosan is allowed" is changed to "There exist some regulations," this element of information has undergone distortion and became less accurate. These content distortions were identified by a coder who was instructed to mark content distortions every time an element of information was transmitted imprecisely from one chain position to the next. We (the authors) independently checked the correctness of the identified distortions. To minimize any ambiguity in the coding process, we provided an illustrated list of five specific types of distortion that need to be tracked in the transcripts:

i) The numerical value has changed or has disappeared. As an example, "The pumping function of the heart decreased by about 25%" becomes "There was myocardial insufficiency."
ii) The qualitative indication of volume, frequency, or probability has changed or has disappeared. As an example, "Germs can possibly become resistant" becomes "Germs develop resistance against it."
iii) An element has moved from a specific to more general class of information. As an example, "It is used for disinfecting hands" becomes "It is a disinfectant." As another example, "She comes from Toronto" becomes "She comes from America."
iv) A previously inexistent element has been added. As an example, "Triclosan accumulates in breast milk" becomes "Children are exposed to triclosan through breastfeeding."
v) The content is obviously wrong. As an example, "He is a cook" becomes "He is a farmer."

Measuring the Message Signal. The risk signal at each transmission (and the change of risk signal between transmissions) refers to the number of positive and negative statements identified in the transcripts. To minimize the subjective interpretation of positive and negative statements, we used two coders who received identical instructions and performed this task independently. Furthermore, we (the authors) independently checked these codings. We then took the mean of the two counts of positive and negative statements (as measured by the coders). The instructions given to both coders included instructive examples of what to measure. The instruction sheet contains the following definitions of statements.

Negative statements. This category corresponds to any sentence that highlights the real or suspected dangers of triclosan and contains one of the following elements:

i) Mention of any negative side effect that triclosan has or can have for people, animals, or the environment. Examples are "It damages the liver," "It is harmful for the environment," or "Mice showed heart problems."
ii) Any expression of a personal feeling or judgment suggesting the danger of triclosan, as well as the report of someone else's negative judgment about it. Example are "You should keep your hands away from it," "I'm scared about it," or "Greenpeace takes a position against it."

The fact that triclosan can be detected in a blood sample or urine sample or that it enters the body through the skin is not considered an explicit mention of danger.

Positive statements. This category corresponds to any sentence suggesting that the use of triclosan is safe or under control, and contains one of the following elements:

i) Any terms mentioning that triclosan is safe or minimizing its danger. Examples are "As long as you don't overdo it, it's not dangerous" or "It's not that bad compared with other things."
ii) Any report of someone else's positive judgment. An example is "The cosmetic industry said such a low dosage is safe."
iii) Any terms suggesting that higher authorities are protecting the consumers. An example is "In Germany, there are strict regulations."

Reliability of measurements. For all transmissions between subjects, the same coding procedure was used independently by both coders. Therefore, the reliability of the results can be assessed by comparing the number of positive and negative statements found by each coder in the same transmission. Overall, we found very close agreement between the coders:

In 56% of the transmissions, both coders reported the same number of positive and negative statements.

In 27% of the transmissions, one coder reported only one more statement than the other.

In 11% of the transmissions, one coder reported two more statement than the other.

Overall, for 94% of the transmissions, the difference between the two coders was two statements or less.

For our analyses, we used the mean number of positive and negative statements, as reported by the two coders.

Simulation Code. The MATLAB (MathWorks) program that has been used to generate the simulation results in the manuscript is available for download in the Harvard Dataverse Network (thedata.harvard.edu/dvn/dv/riskcommunication).

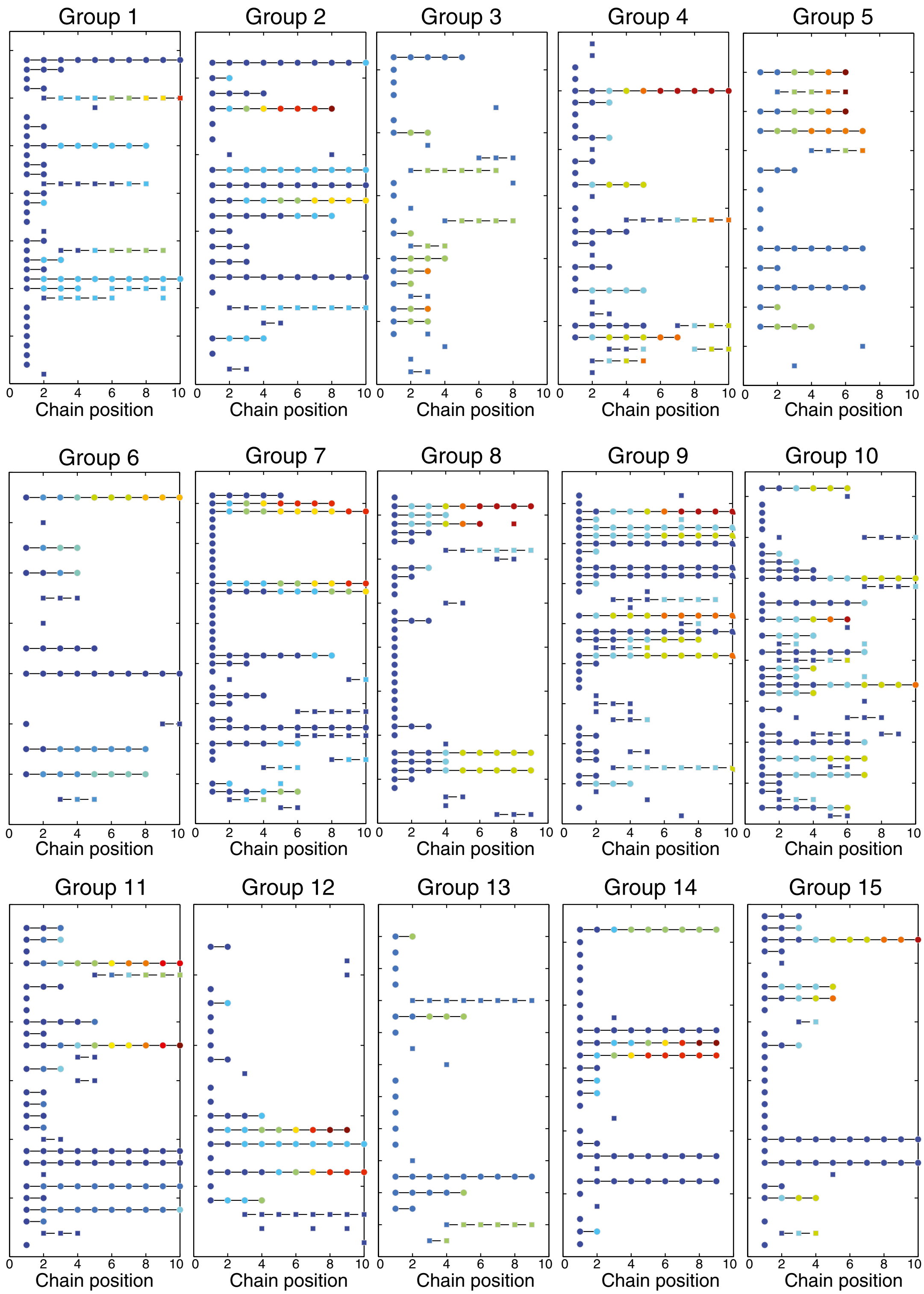

Fig. S1. Topological maps of information propagation for all 15 replications of the experiment. Group 8 is shown and commented on in the main text (Fig. 1).

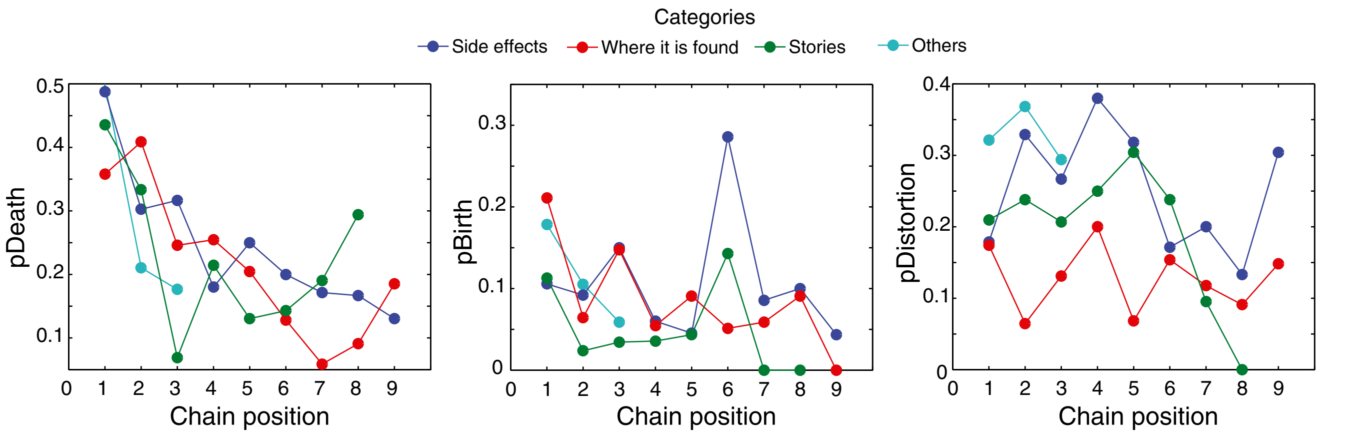

Fig. S2. Hazard functions showing the probabilities that a piece of information disappears (*Left*), appears (*Middle*), or gets distorted (*Right*) at each chain position, compared between types of information. The four lines correspond to four different categories of information indicated at the top of the figure. No specific difference is found between categories, suggesting that any pieces of information can disappear (pDeath), appear (pBirth), or get distorted (pDistortion), regardless of the nature of the information that is communicated.

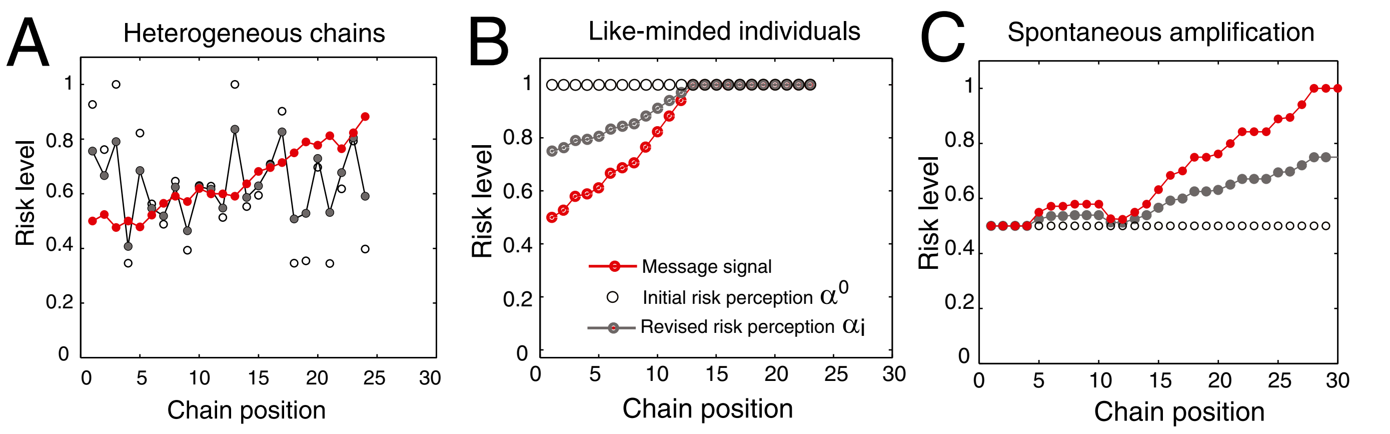

Fig. S3. Representative examples of model simulations. (*A*) People composing the chain express different initial opinions. Heterogeneity of initial judgments is based on the values observed in our experiment (average value = 0.65 and SD = 0.22). As observed in our data, the model predicts a gradual amplification of the risk signal (red curve) and a considerable variability of revised judgments, which reflect the initial diversity of risk perceptions. (*B*) In chains of like-minded people, the amplification process is exacerbated. As the message is communicated from one person to another, it gradually mutates to fit the initial view of the population. At the end of the chain, individuals are no longer influenced by the injected message. (*C*) Illustration of a spontaneous amplification case, where a neutral message injected in a neutral chain generates high levels of risk perceptions. This phenomenon results from the amplification of random mutations of the signal. In all three figures, the injected message is neutral with $n^+ = 10$ positive statements, and $n^- = 10$ negative statements (and thus signal $\sigma = 0.5$).

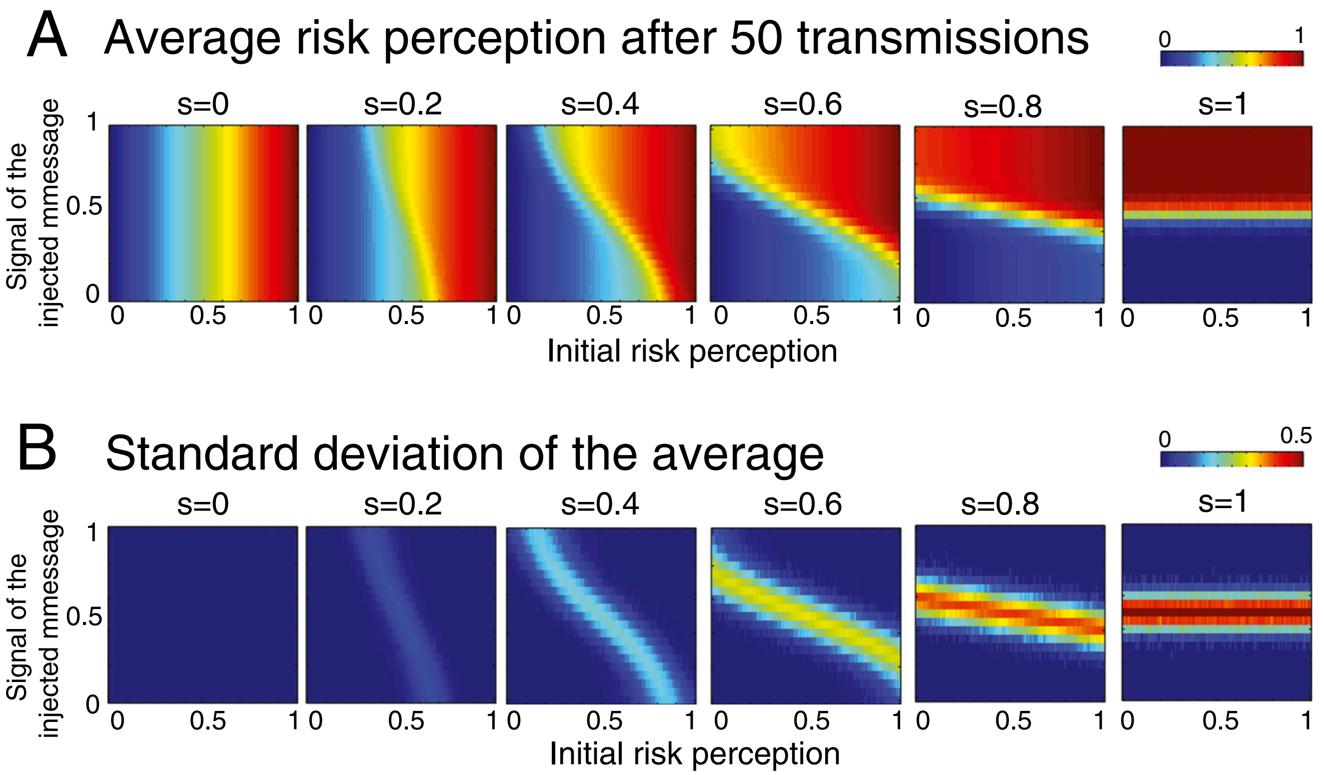

Fig. S4. Effects of varying the social influence coefficient s on the model's predictions. At $s = 0$, social influence is inexistent: The final opinion of the individuals equals its initial value (*A*, *Left*), and the system is deterministic (*B*, *Left*). (*A*) As s increases, the S-shaped transition zone emerges and becomes gradually sharper. (*B*) At the transition zone, the system becomes increasingly unpredictable. (*A*, *Right*) At $s = 1$, social influence is at the maximum level, and individuals fully adhere to the risk message regardless of their initial opinion. This latter case exhibits strong amplification effects, where any message with a signal lower than 0.5 converges to 0. Conversely, a message with a signal higher than 0.5 converges to 1. (*B*, *Right*) Neutral messages carrying a signal around 0.5 can equally tip up to 1 or down to 0.

Other Supporting Information Files

SI Appendix (PDF)