# Supporting Online Materials
# An experimental study of inequality and unpredictability in an artificial cultural market

Matthew J. Salganik Peter Sheridan Dodds Duncan J. Watts

## Experimental design

As stated in the main text, subjects entering the experiment were randomly assigned into either the *independent* condition or the *social influence* condition. Subjects in the independent condition had no information about the previous behavior of others and so were forced to make their decisions about the songs independently. However, subjects in the social influence condition were given information about the behavior of others which they could use, or ignore, when making their decisions. Any difference in success outcomes for the songs between these two groups can be attributed to presence of social influence.

Our design also had an additional step. In order to better understand unpredictability, subjects in the social influence condition were further randomly assigned into one of eight influence "worlds". Each subject was given information only about the behavior of others in their influence world. We thus created multiple "histories" to determine to what extent indistinguishable groups of subjects, starting at the same initial condition, and choosing from the same set of songs can generate different success outcomes. We only needed one independent condition realization because the behaviors of the subjects in this condition were independent. A schematic of the experimental design is shown in Fig. S1.

The assignment of subjects was done such that 20% of the subjects were assigned the independent condition and 10% were assigned to each of the eight social influence worlds. For each experiments, this allocation resulted in about 700 subjects in each of the social influence worlds and about 1,400 in the independent condition. The reason for this allocation scheme will become clear when we discuss our measure of unpredictability.

## Subject experience during the experiment

The entire framework of the experimental design was unknown to the subjects. Upon entering the website (`http://musiclab.columbia.edu`) subjects were presented with a welcome screen telling them that they

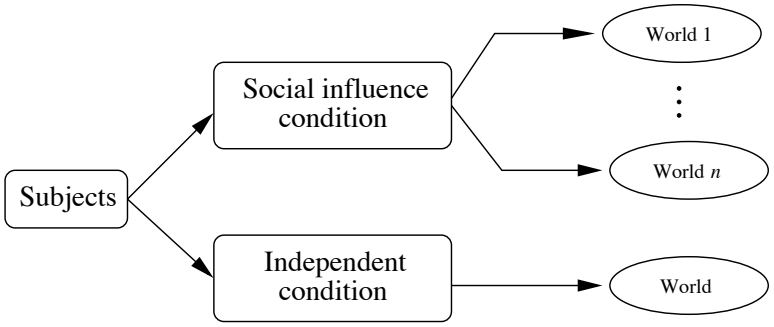



Figure S1: Schematic of the experimental design.

were about to participate in a study about musical tastes and that in exchange for participating they would be offered the chance to download some free songs by up-and-coming artists. Subjects next gave their informed consent, filled out a brief survey, and were shown a page of instructions. Finally, subjects were presented with a menu of 48 songs.

In experiment 1, the songs were presented in a three column jukebox-type design (see Fig. S2) and displayed in a random order to each subject. By randomizing the order for each subject we avoided favoring any songs by placing them in advantageous screen-locations. However, the specific order for each subject was fixed for the entire experiment. Subjects in the social influence condition were also presented with the song download counts in their world while subjects in the independent condition were not.

In experiment 2, the songs were presented in a one column design (see Fig. S3). Subjects in the social influence worlds were presented the songs *sorted by number of downloads*, along with the download counts in their world. If several songs shared the same number of downloads, the ordering of the songs was determined randomly for each user. Subjects in the independent condition in experiment 2 were presented with the songs in the same one column design, but in random order and without the download counts.

Once at the menu of songs, if a subject clicked on a specific song, they were taken to a new screen where the song automatically began playing in a Macromedia Flash Player, streamed in the mp3 format encoded at 96kbps (Fig. S4). While a subject listened to the song they were asked to rate it on a scale from 1 to 5 stars which could be done at any time while the song was playing; subjects did not need to wait for the song to complete. After the rating was recorded, subjects were asked if they would like to download the song (Fig. S5). After making the download decision, subjects were returned to the menu of 48 songs and were able to choose again.

Once a subject had listened to as many songs as they wished, they could click "log off" and were taken to a screen thanking them for participating and providing them links to the webpages of all 48 bands. Subjects who returned to the website while the experiment they participated in was still underway were automatically


Figure S2: Screenshot of the song menu in the social influence world in experiments 1. Screenshot from the independent condition (not shown) was identical except that the download counts to the right of each song are removed.

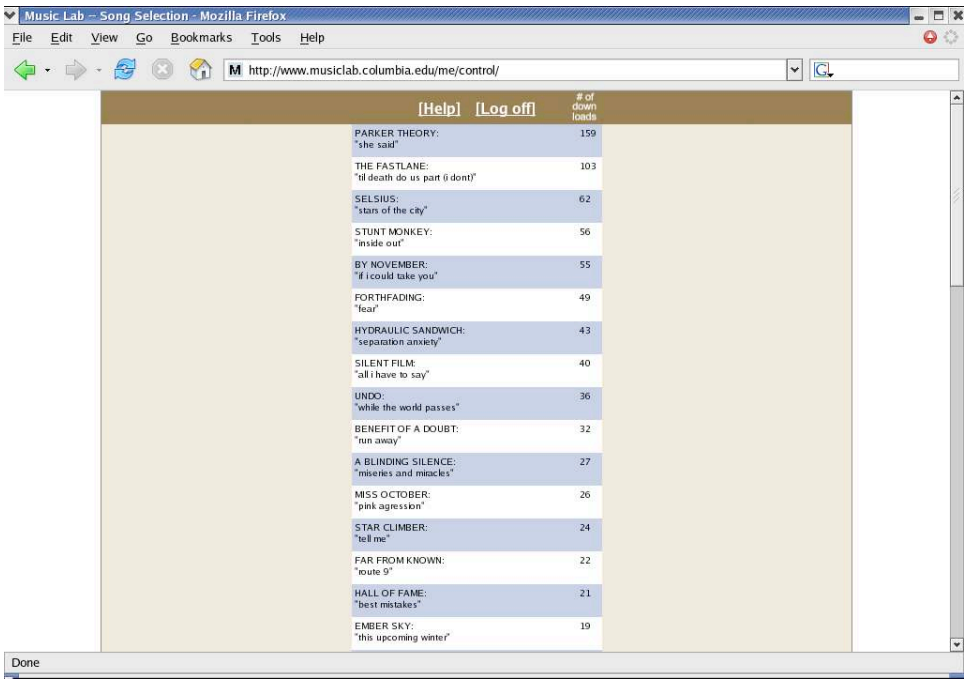



Figure S3: Screenshot of the song menu in the social influence world in experiments 2. Screenshot from the independent condition (not shown) was identical except that the download counts to the right of each song are removed.

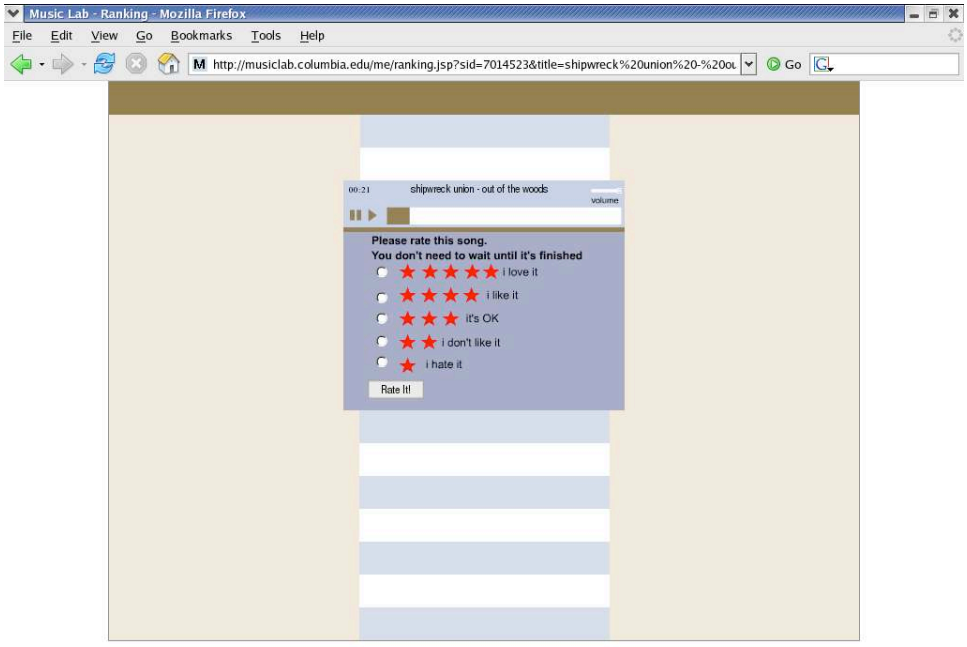



Figure S4: Screenshot of the listening screen. While a song was playing subjects where required to rate it on a scale of 1 to 5 stars. This rating could be submitted before the song was finished playing.

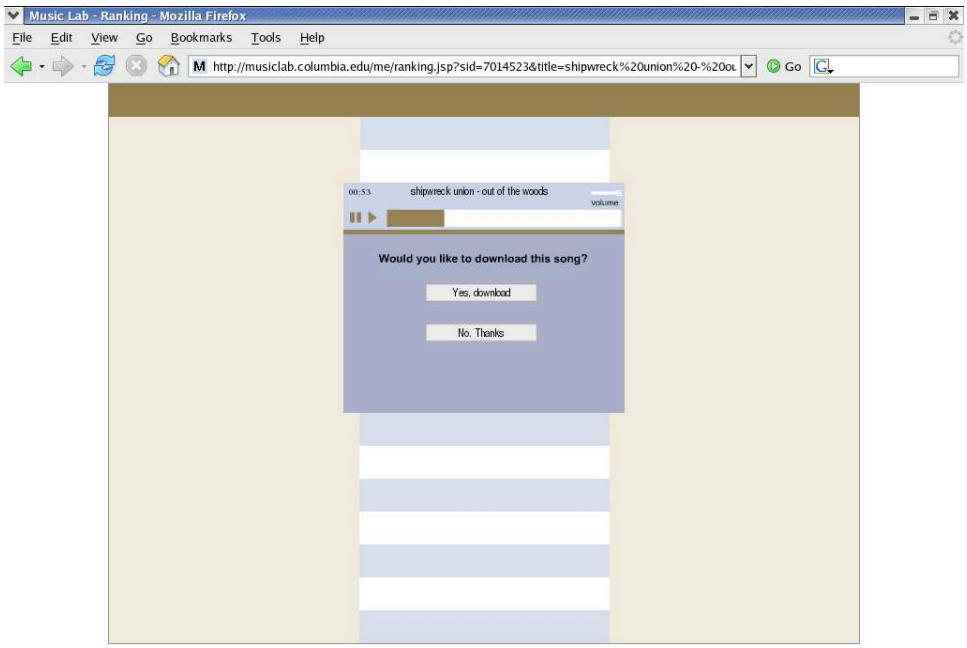



Figure S5: Screenshot of the download decision screen. After rating the song, subjects had to decide to download the song or not.

| Category | Experiment 1 ($n = 7,209$) (% of participants) | Experiment 2 ($n = 7,207$) (% of participants) |
|---|---|---|
| Female | 36.4 | 73.9 |
| Broadband connection | 74.1 | 69.0 |
| Has downloaded music from other sites | 60.4 | 62.4 |
| Country of Residence | | |
| United States | 79.8 | 81.8 |
| Canada | 4.5 | 4.4 |
| United Kingdom | 4.4 | 4.7 |
| Other | 11.3 | 9.1 |
| Age | | |
| 14 and younger | 11.5 | 16.0 |
| 15 to 17 | 27.8 | 34.9 |
| 18 to 24 | 38.5 | 39.2 |
| 25 and older | 22.3 | 9.9 |

Table S1: Descriptive statistics about the subjects.

| | Experiment 1 | | | Experiment 2 | | |
|---|---|---|---|---|---|---|
| | Influence ($n = 5,760$) | Independent ($n = 1,449$) | Total ($n = 7,209$) | Influence ($n = 5,757$) | Independent ($n = 1,450$) | Total ($n = 7,207$) |
| Number of listens | 21,263 | 5,195 | 26,558 | 19,557 | 5,454 | 25,011 |
| Mean per subject | 3.7 | 3.6 | 3.7 | 3.4 | 3.8 | 3.5 |
| Median per subject | 1 | 1 | 1 | 1 | 1 | 1 |
| Number of downloads | 6,626 | 1,578 | 8,203 | 8,106 | 2,192 | 10,298 |
| Mean per subject | 1.2 | 1.1 | 1.1 | 1.4 | 1.5 | 1.4 |
| Median per subject | 0 | 0 | 0 | 0 | 0 | 0 |

Table S2: Descriptive statistics on subject behavior in the two conditions and overall.

returned to their world and taken to the appropriate song menu without the need to re-register. Subjects from experiment 1 who returned to the website during experiment 2 were prevented from participating.

## Subject recruitment

Experiment 1 took place from October 7, 2004 to December 15, 2004 (69 days) and involved 7,209 subjects. Immediately after completing experiment 1, we began experiment 2 which ran from December 15, 2004 to March 8, 2005 (83 days) and involved 7,207 subjects. Most subjects were recruited from `http://www.bolt.com`, a website popular with teens and young adults from the United States. Demographics about these subjects are presented in Table S1 and summary statistics about their behavior is presented in Table S2.

We note that there was a change in percentage of females from experiment 1 to experiment 2. Subjects in both experiments were drawn from `http://www.bolt.com`, but they were drawn from different parts of the website. A majority of the subjects in experiment 1 were likely drawn from the "music" and "free-


Figure S6: Banner used to recruit subjects from `http://www.bolt.com` for experiment 2.

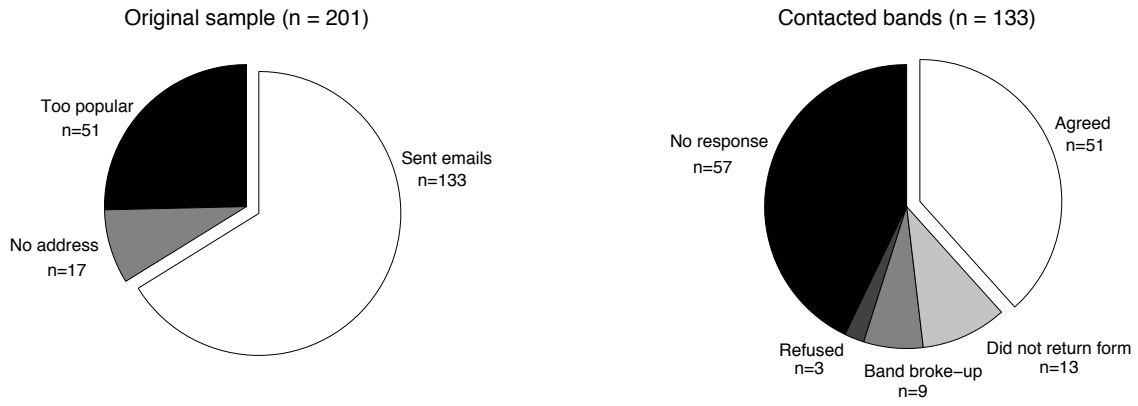
stuff" sections while a majority of the subjects in experiment 2 were likely drawn from a special email sent to a set of Bolt users and from banner ads in all sections of the site (for example, Fig. S6). Another potential reason for the difference is that while experiment 1 was underway, the project was mentioned on the popular blog `http://www.kottke.org` which probably has an older, more male readership. Ideally these differences in recruitment between experiments would not have occurred, but we do not believe that they had a substantial effect on our findings.

## Music selection

The music for the experiment comes from `http://www.purevolume.com`, a website where bands can create homepages and post their music for download. In July 2003 there were approximately 42,000 bands with homepages. Preliminary research revealed that the quality of the music of these bands was extremely variable with a large number having very poor audio quality. However, `http://www.purevolume.com` also hosted of a set of premium member bands who paid approximately $10 per month for additional features on their homepages. There were approximately 1,000 premium bands, and we took a random sample from these bands.

Initially, about 200 bands were selected. The experiments required bands that are unknown to the subjects so we screened out any band that had played in more than 10 states, or had played more than 15 concerts in the past 30 days, or had appeared on the Warped Tour, or had 30,000 or more hits on their purevolume page. These screening criteria are ultimately arbitrary, but they are reasonable. We have no reason to believe that the results would be any different if other reasonable criteria were used. In all, these criteria removed 51 bands. In addition, 17 bands could not be contacted because they did not have a publicly available email address. The remaining 133 bands were contacted via email (results summarized in Fig. S7A). In order to minimize non-response bias, all non-responding bands received two follow-up emails spaced at one week intervals. In the end, 51 of these bands agreed to be in the study and provided us with a song of their choice, the other bands becoming ineligible for a variety of reasons (results summarized in Fig. S7B).

Preliminary pilot testing revealed that, for the song menu used in experiment 1 (Fig. S2), the maximum number of songs that could be legibly presented on a typical computer screen was 48. Thus, we took a




(a) Original sample (n = 201)

(b) Contacted bands (n = 133)

Figure S7: Pie charts showing various aspects of attrition for the sample of bands selected from the music website `http://www.purevolume.com`. Approximately, 40% of the contacted bands agreed to be in the study.

sample of 48 of the 51 bands to be in the experiments. A list of these bands and songs can be found in table S3.

In order to check that our initial screening criteria filtered out music that might be known to the subjects, we presented the list of bands and songs to two different experts in popular music: a DJ at the Barnard College student radio station and the music editor for `http://www.bolt.com`. Neither expert recognized any of the bands or songs. As an additional test, we surveyed subjects about their familiarity with the 3 bands who agreed to participate, but were ultimately not included because we were limited to 48 bands. We chose to ask only about the bands that were ultimately not included because having the same bands in the survey and experiment might have biased subjects' music preferences.

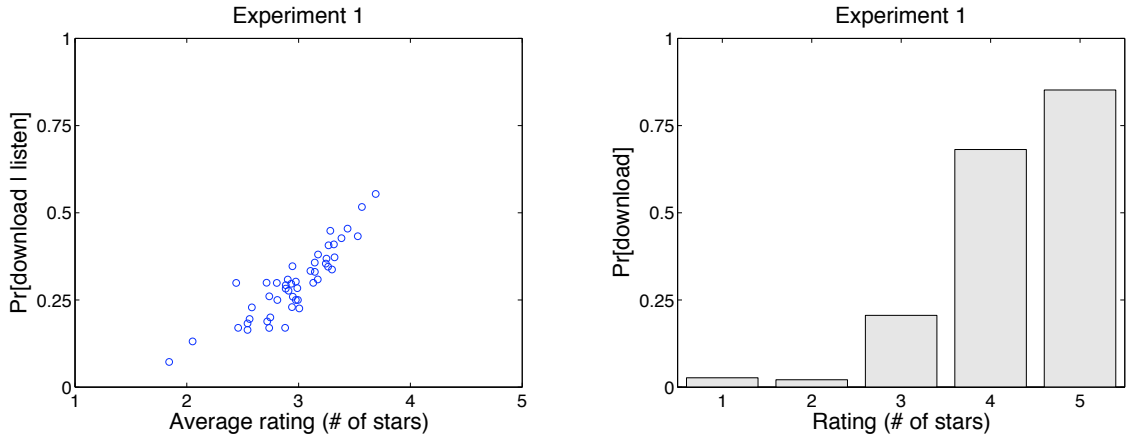
In table S4 we compare the subjects' familiarity with three bands from our pool of potential bands to one fake band. The data suggest that there was a large amount of social desirability bias in responses — 14% of subjects reported hearing of the fake band Peter on Fire and 2% reported being familar with their music. The responses for the fake band are very similar to the responses for the real bands. The high recognition rate for the band Remnant Soldier is probably a question ordering effect; this question was asked immediately after a question about familiarity with the very popular band U2. In future studies we recommend randomization to avoid this problem. Taken together, these results, along with our screening, lead us to believe that the music used in the experiment was essentially unknown. Also, while the experiments were in progress, we monitored the success of the bands and found nothing which would lead us to believe that there were any significant changes.

| Band name | Song name |
|---|---|
| 52metro | Lockdown |
| A Blinding Silence | Miseries and Miracles |
| Art of Kanly | Seductive Intro, Melodic Breakdown |
| Beerbong | Father to Son |
| Benefit of a Doubt | Run Away |
| By November | If I Could Take You |
| Cape Renewal | Baseball Warlock v1 |
| Dante | Life's Mystery |
| Deep Enough to Die | For the Sky |
| Drawn in the Sky | Tap the Ride |
| Ember Sky | This Upcoming Winter |
| Evan Gold | Robert Downey Jr. |
| Fading Through | Wish me Luck |
| Far from Known | Route 9 |
| Forthfading | Fear |
| Go Mordecai | It Does What Its Told |
| Hall of Fame | Best Mistakes |
| Hartsfield | Enough is Enough |
| Hydraulic Sandwich | Separation Anxiety |
| Miss October | Pink Aggression |
| Moral Hazard | Waste of my Life |
| Nooner at Nine | Walk Away |
| Not for Scholars | As Seasons Change |
| Parker Theory | She Said |
| Post Break Tragedy | Florence |
| Ryan Essmaker | Detour_(Be Still) |
| Salute the Dawn | I am Error |
| Secretary | Keep Your Eyes on the Ballistics |
| Selsius | Stars of the City |
| Shipwreck Union | Out of the Woods |
| Sibrian | Eye Patch |
| Silent Film | All I have to Say |
| Silverfox | Gnaw |
| Simply Waiting | Went with the Count |
| Star Climber | Tell Me |
| Stranger | One Drop |
| Stunt Monkey | Inside Out |
| Sum Rana | The Bolshevik Boogie |
| Summerswasted | A Plan Behind Destruction |
| The Broken Promise | The End in Friend |
| The Calefaction | Trapped in an Orange Peel |
| The Fastlane | Til Death do us Part (I don't) |
| The Thrift Syndicate | 2003 a Tragedy |
| This New Dawn | The Belief Above the Answer |
| Undo | While the World Passes |
| Unknown Citizens | Falling Over |
| Up Falls Down | A Brighter Burning Star |
| Up for Nothing | In Sight Of |

Table S3: List of the 48 bands used in the experiment. These bands were randomly selected from the website `http://www.purevolume.com`. Several tests were conducted which allow us to conclude that these bands were essentially unknown.

| | % heard of | % familiar with |
|---|---|---|
| Real Bands | | |
| Guys on Couch | 11.0 | 1.1 |
| Grover Dill | 10.5 | 1.1 |
| Remnant Soldier | 19.9 | 2.9 |
| Fake Band | | |
| Peter on Fire | 13.7 | 1.8 |

Table S4: Investigating the popularity of the potential bands from our sample in comparison to a fake band. Subjects in experiment 1 and 2 reported being about as familiar with an fake band (Peter on Fire) as three potential bands from our sample. The high recognition rate for Remnant Soldier is likely a question ordering effect — it was asked immediately after the well known band U2.




(a) Comparing behavioral and self-report data

(b) Comparing behavioral and self-report data

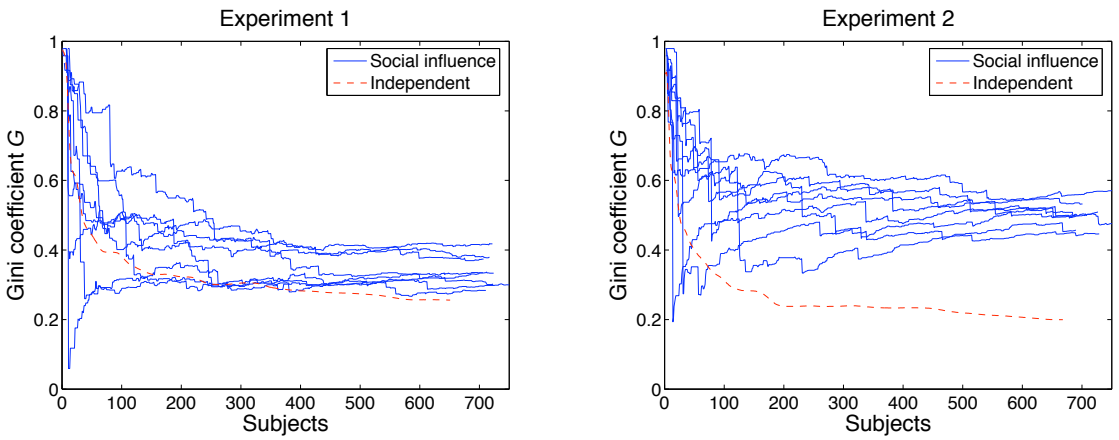
Figure S8: Plots comparing the download decisions to the rating decisions. These results suggest that the two measures are consistent. Results from experiment 2 (not shown) were essentially the same.

## Data analysis

We measured success based on the market share of downloads that belonged to a specific song. The market share, $m_i$ of song $i$ is defined as,

$$m_i = \frac{d_i}{\sum_{k=1}^{S} d_k} \tag{1}$$

where $d_i$ is the number of downloads for song $i$ and $S$ is the number of songs. This definition of success is based on the subjects' behavior, rather than their self-reported liking of the songs, as measured by their ratings from 1 to 5 stars. As a check, we compared these two measures and found them to be consistent. In Fig. S8A we see that songs which received higher average ratings (measured in stars) had higher probabilities that a listen would result in a download ($r = 0.89$). In Fig. S8B we see that the higher rating a subject gave a song, the more likely that the subject downloaded the song. Results from experiment 2 are essentially identical (results not shown). Overall, the similarity between these two measures gives us confidence that our behavioral measure is meaningful.




(a) Gini coefficient, experiment 1

(b) Gini coefficient, experiment 2

Figure S9: Dynamics of the Gini coefficient $G$ in experiment 1 and 2. The final values of each trajectory are the values reported in the Fig. 1.

Given that we use the market share of songs as a measure of success, we measure inequality of success with one of the most common metrics, the Gini coefficient, $G$, which is defined as follows,

$$G = \frac{\frac{1}{S^2}\sum_{i=1}^{S}\sum_{j=1}^{S}|m_i - m_j|}{2 \cdot \frac{\sum_{k=1}^{S} m_k}{S}}. \tag{2}$$

The Gini coefficient can be interpreted as the expected difference in market share between two randomly chosen songs scaled so that it falls between 0 and 1 with 0 representing complete equality and 1 representing complete inequality. As stated previously, the independent condition has twice the number of subjects ($n \approx 1,400$) as each social influence world ($n \approx 700$), for reasons that will be clear when we present our measure of unpredictability. In order to ensure that our comparison between the two conditions was based on the same number of subjects, we randomly split the independent condition into two groups and then calculated the Gini coefficient for one of these groups. We repeated this splitting procedure 1,000 times and produced a distribution of replicate values of $G$. The value of $G$ reported in Fig. 1 for the independent condition is the mean of these 1,000 replicate values. Also, we used the distribution of replicate values to conduct a test of statistical significance. The difference between a randomly chosen Gini coefficient from one of the eight influence worlds and randomly chosen replicate Gini coefficient from the independent world was less than 0 with $p < 0.01$ in experiment 1 and $p < 0.001$ in experiment 2. Thus, the difference in observed Gini coefficients between the two conditions is statistical significant in both experiments. Finally, we can examine the dynamics of the Gini coefficient as the experiment progresses (Fig. S9). The final values of each trajectory are the values reported in Fig. 1. The Gini coefficients were relatively stable indicating that we probably would not have observed substantially different results with more subjects.

In addition to the Gini coefficient, we measured inequality using two other common measures, the coefficient of variation and the Herfindahl index. The results were qualitatively unchanged. We could not consider any of the logarithm-based measures, standard deviation of the logarithms and Theil entropy, because these measures are not defined in cases where a song has 0 downloads which occurred in one of the social influence worlds in experiment 1. For more on all of these measures of inequality, see Coulter (1989).

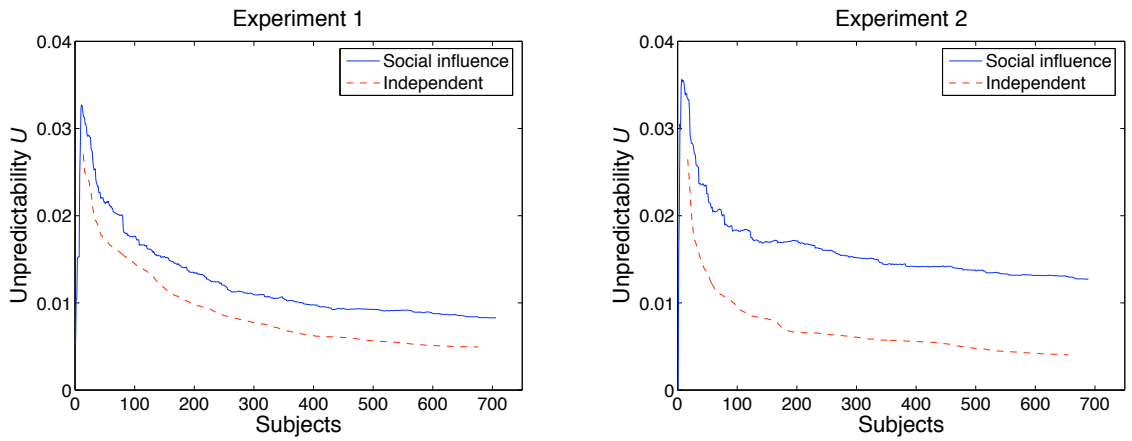
To measure unpredictability we examined the variation in success of a song across realizations. If a song had the same outcome in all realizations then its unpredictability was 0. However, if the outcomes varied across realizations, then there was an inherent unpredictability in the success of the song. We defined $u_i$ as a measure of the unpredictability of song $i$ to be the average difference in market share across all possible pairs of realizations. That is,

$$u_i = \frac{\sum_{j=1}^{R} \sum_{k=j+1}^{R} \mid m_{i,j} - m_{i,k} \mid}{\binom{R}{2}}, \tag{3}$$

where $m_{i,j}$ is the market share for song $i$ in realization $j$, and $\binom{R}{2}$ is the number of pairs of realizations. The unpredictability, $U$, for an experimental condition is then the average of the unpredictability of the songs in that condition,

$$U = \frac{\sum_{i=1}^{S} u_i}{S}. \tag{4}$$

In the independent condition we have only one realization, but, as noted previously, it has twice as many subjects as each social influence world. Thus, for the independent condition, we randomly split the subjects into two independent realizations and calculated $u_i$ and $U$ with these two realizations. We repeated this splitting procedure 1,000 times and produced a distribution of replicate values of $U$. The value of $U$ reported in Fig. 2 is the mean of this distribution. To calculate a measure of statistical significance we compared the distribution of replicate values from the independent condition to the distribution of calculated $U$ values for the 28 ($\frac{8\times 7}{2}$) possible pairs of influence worlds. The difference between a replicate influence world and a replicate independent world was less than 0 with a probability of $p < 0.01$ in experiment 1 and $p < 0.001$ in experiment 2. Thus, the difference in unpredictability across conditions is statistically significant. Finally, we can examine the dynamics of the unpredictability $U$ as the experiment progresses (Fig. S10). The final values of each trajectory are the values reported in Fig. 2. As with our measure of inequality, the unpredictability was relatively stable indicating that we probably would not have observed substantially different results with more subjects.




(a) Unpredictability, experiment 1

(b) Unpredictability, experiment 2

Figure S10: Dynamics of unpredictability $U$ in experiment 1 and 2. The final values of each trajectory are the values reported in Fig. 2.

## Measures to ensure data quality

In all experiments researchers must take steps to ensure that data are generated by the appropriate set of subjects in situations that match the experimental design and that the subjects have no malicious intent. These problems can be more difficult to deal with in web-based experiments where researchers have less control over subject recruitment and behavior than they would have in a standard laboratory-based experiment. Because of this limited control, some of the data from our experiments are possibly unsound. Instead of preventing this unsound data generation, and hence giving subjects incentive to provide us with false information, we allowed all subjects to participate in all situations, but flagged data that could have been unsound and excluded them from our analysis.

For example, our experimental design required that a subject's information about the behavior of others be limited to what we provided them (or did not provided them). Information contamination leading to unsound data could have occurred a number of ways: 1) between two subjects from different two influence worlds 2) between two subjects from the independent condition and 3) between a subject in the independent condition and a subject in an influence world. Unlike in a traditional laboratory-based experiment, we were not able to physically isolate the subjects to prevent this information contamination. As such, we flagged for exclusion data generated in several cases where the subject behavior could have possibly been influenced by information that was outside of the experimental design.

The first step in this data-flagging process was based on a survey that all subjects completed. On this survey subjects were asked to select, from a list of choices, all of the ways that they heard about the experiment. If a subject reported "friend told me about a specific song" or "friend told me about a specific

band" all data generated by that subject were flagged. However, data generated by subjects who reported "friend told me about the experiment in general" were not flagged. We also flagged all data generated after either the subject clicked "log-off" or 2 hours had passed since the subject registered. These data were flagged in order to exclude data where the subject could have participated, discussed the music with friends, and then returned with outside information.

In addition, to prevent information contamination within and between experiments, we placed several cookies — small pieces of information — into the subject's web browser. These cookies ensured that if a subject returned to the experiment, the subject would be placed in the same condition and same world without having to re-complete the registration process. The cookies also limited the possibility of subjects from experiment 1 participating in experiment 2.

Our flagging criteria were quite strict and so we probably flagged data which was not contaminated. However, we cannot rule out the possibility that some contaminated data was not flagged. Any information contamination across influence worlds would have likely had the effect of decreasing the differences across worlds and thus decreasing our unpredictability measure. Information contamination within the independent condition would have likely increased the inequality in the independent condition. Finally, information contamination between a social influence world and the independent condition would likely increase the correlation between quality and success. Thus, our findings on inequality, unpredictability, and the relationship between quality and success represent a lower-bound on the possible values that could have occurred in a perfectly clean experiment.

In addition to problems with the isolation of subjects, when doing a web-based experiment, or any other experiment, one has to take a number of steps to guard against the possibility of malicious subjects who intend to disrupt the experiment. This problem, while not limited to web-based experiments, is perhaps a larger issue in this set of experiments than in most. For example, members of one of the bands might have tried to artificially inflate the download count of their song. To prevent this possibility, each subject was allowed to download a specific song as many times as they liked, but could only add one to the displayed download count for that song. Members of the bands might have also tried to manipulate the results by sending their fans to the experiment. As such, we flagged all data generated by people who reported on our survey that they heard about the experiment from "one of the bands." We also checked our web-server log to ensure that we were not receiving subjects from the websites of any of the bands. In two cases, links to the experiment was posted on bands' websites, but these links were detected quickly and both bands complied with our email request to remove the link.

An additional class of malicious subjects could have simply wished to disrupt the experiment for no specific reason. To prevent against these subjects, the experiment was run appropriate security precautions

using the latest software (Apache 2.0, MySQL 4.0, and Tomcat 5.0) with strict firewall settings.

Despite all of our security precautions it was still possible for a subject to manipulate our results. For example, there is no way that we could prevent the same person from registering from several different computers and providing us with false information each time. However, given that subjects have little incentive to undertake this behavior, we think that this probably did not occur. Taken together our data-quality measures give us confidence that our data are reasonably clean. Of course we cannot rule out all possible problems, but we have not seen any patterns in the data that indicate data contamination or malicious manipulation occurred.

## Robustness of results to specific design choices

These two experiments represent only a small portion of the parameter space of all possible experiments using this design. For example, system parameters like the strength and type of social signal, the subject population, the distribution of quality of the songs, and the number of songs probably influence the magnitude of the observed outcomes. Based on our experience with these experiments, we offer a few predictions.

We suspect that other methods of strengthening the social signal would increase the inequality and unpredictability. For example, in our experiment we chose to present the number of previous downloads, the band name, and song name all in the same size font. If, for example, we had presented the download counts in a larger font we suspect that the inequality and unpredictability would be greater.

However, other methods of changing the social signal may have ambiguous effects on outcomes. For example, in our experiments, the social signal was anonymous, in the sense that subjects did not have any information about the characteristics and behavior of previous subjects. If the social signal was instead somehow linked to the identities of the previous subjects, one could imagine that since subjects may be more strongly influenced by "people like them," the cumulative advantage process could be weakened or strengthened depending on the distribution of subjects' identities.

Given the type of signal that we chose to use, we suspect that the process of social influence observed in the experiments is relatively general, but may be more pronounced with our subject pool (teenagers from the U.S.). We suspect that if the experiment was re-run using a different subject pool that different songs would become successful, but that the overall amount of inequality and unpredictability would be similar. Further empirical work in this area is needed.

Switching from characteristics of the subjects to characteristics of the songs, we expect that if the songs were more similar in quality, then the inequality in success would be less, but the unpredictability would be greater. Recall, that in these experiments we did not directly set the distribution of quality; rather, it was

determined by the songs on `http://www.purevolume.com`.

Another key system parameter for the songs is the number used. Because choice overload is so pervasive in cultural markets, we chose to use 48 songs in the experiments — the maximum that could fit on a computer screen when presented with the song menu used in experiment 1 (Fig. S2). We conjecture that if we had used more songs, the observed inequality and unpredictability would increase. Whatever the final number of songs used, it is likely important that this number is much larger than the number of songs that each subject listens to.

These speculations are suggestive, and clearly more research is needed. However, the speculations do suggest that the qualitative findings of the experiments are likely robust to reasonable design choices.

## References

Coulter, P. B. (1989). *Measuring Inequality: A Methodological Handbook*. Westview Press, Boulder.