

Using Music-Evoked Affect to Enhance Performance on Micro-Task Crowdsourcing Platforms

Rob Morris

MIT Media Lab
Cambridge, MA 02139
r Morris@media.mit.edu

Mira Dontcheva

Advanced Technology Labs
San Francisco, CA 94103
mirad@adobe.com

Elizabeth Gerber

Northwestern
University
Evanston, IL 60208
egerber@northwestern
.edu

Adam Finkelstein

Princeton University
Princeton, NJ 08542
af@cs.princeton.edu

ABSTRACT

Crowdsourcing is an emerging paradigm used to harness the untapped cognitive abilities of thousands of individuals. While desirable for its speed and low-cost, performance in crowdsourcing environments varies greatly, especially for tasks requiring extended cognitive effort or creative insight. To improve performance on crowdsourcing platforms, we propose embedding affective primes within online digital interfaces. There is a well-established link between positive affect and performance, but applied affective priming techniques have yet to be explored in a crowdsourcing environment. In two experiments, we explore how these techniques can be deployed in crowdsourcing environments to improve performance. We find that positive musical primes, played immediately prior to a creative problem-solving task, enhance performance relative to a neutral, non-priming condition. This study is the first to report significant affective priming effects in a real-world human computer interaction context. We discuss practical implications for task design in crowdsourcing and features of designs that may induce affect.

Author Keywords

crowdsourcing, performance, affective computing, problem-solving, computational primes, music

INTRODUCTION

Crowdsourcing is an emerging paradigm used by industry and governmental agencies to harness the unique cognitive abilities of thousands of individuals [1]. The vision is to harness the power of the crowd to address challenges ranging from improving online search to finding cures to diseases, often with the goal of addressing problems that are difficult or impossible to solve via completely automatic processes. To realize this vision, organizations distribute

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Segal Design Institute, McCormick School of Engineering, Northwestern University, Evanston, IL, USA January 18, 2012, Tech Report #002



Our studies show that music-evoked affect can improve insight problem-solving performance in crowdsourcing environments.

tasks to the crowd ranging from simple image tagging to more complex activities such as protein folding [2] or furniture design [3].

Micro-task platforms such as Amazon's Mechanical Turk have typically been used to distribute simple cognitive tasks. Unlike peer-production platforms such as Wikipedia, workers are not incentivized to spend extended cognitive resources to complete a task, but rather minimal effort and resources to complete many small tasks quickly in order to be paid [4].

While micro-task platforms are desirable for their speed, scalability, and low cost, task performance varies greatly, especially for tasks requiring cognitive effort [5], thus limiting the types of tasks that can be successfully distributed [6]. To increase the types of tasks that can be distributed in micro-task environments, human-computer interaction scholars are investigating approaches to improve performance. Scholars find that strategies such as administering qualification pre-tests, incorporating verification tasks [4], decomposing complex tasks into small, simple subtasks [6], and aggregating work across many workers [2] can help improve task performance. Conversely, performance-contingent strategies such as

increasing financial incentives [5] are not as effective at improving task performance. While this research has inspired new best practices for task design, human-computer interaction (HCI) researchers have yet to take an *affective* approach to improving performance, despite a demonstrated link between positive affect and performance on problem-solving tasks [8].

In this paper, we present two studies that use music as an affective prime to enhance creative problem-solving on Amazon's Mechanical Turk. In the first study, we examine the effects of listening to music immediately prior to a task, as part of a supposed audio-based verification task. In this context, we find that listening to "positive" music increases subsequent on-task performance relative to a neutral verification task condition. In a second study, participants hear background music while completing the problem-solving task. In this context, we find that affect-laden music has no significant impact on performance. We discuss task design implications based on these findings.

This research is the first to demonstrate that music-based affective primes can improve task performance in the context of online crowdsourcing. We illustrate how priming techniques previously studied in the laboratory can be drastically shortened and can be implemented in a real-world micro-task crowdsourcing platform. We present this research in the context of what we call *affective crowdsourcing*, the investigation of affect as it relates to crowdsourcing.

Crowdsourcing

Online crowdsourcing allows people to distribute work to a network of individuals with Internet access. Individuals in a crowd may be tightly coupled, knowing and trusting each other, or loosely coupled, anonymously working on related tasks with limited knowledge of each other and the requester. In a given crowd, workers may be motivated to participate for intrinsic and/or extrinsic reasons [9]. When intrinsically motivated, behavior is driven by interest or enjoyment in the task at hand [10]. For example, crowd workers may enjoy labeling images as part of the ESP game [11] or find researching and writing articles for Wikipedia useful for realizing their potential as a writer. When extrinsically motivated, people are incentivized by instrumental gains, such as monetary or social gains, that are separate from the task at hand [10]. For example, crowd workers may contribute to open software projects to obtain the respect of their peers or they may write article reviews on Amazon's Mechanical Turk in exchange for pay.

Motivation for Working

While people can be motivated to engage in crowdsourcing for both intrinsic and extrinsic reasons, researchers find that when extrinsic motivations (such as monetary incentives) are offered, motivation to complete a task often declines [12]. For-pay micro-task markets such as Mechanical Turk primarily motivate workers with monetary compensation. On such platforms, requesters, or people who create,

distribute, and pay for the completed tasks, create Human Intelligence Tasks or HITs which specify the task, requirements, and compensation amount. Tasks range from simple tasks such as cleaning data to more complex tasks such as creating and moderating. As an example, companies such as AOL and Casting Words hire workers to categorize content on websites, transcribe audio files, and edit content [13]. Tasks typically take a few seconds to a few minutes to complete. Workers login to the online marketplace, preview the available tasks, and chose a HIT they wish to perform. Amazon's Mechanical Turk is a particularly popular platform, reaching 500,000 workers and offering over 180,000 HITs at any one time [13].

As of this writing, Mechanical Turk (MTurk) workers tend to fall into one of two primary groups: young, college-educated, middle income, American females or young, college-educated, low income, Indian males [14]. Given that most workers have a college education, we can assume a basic competence in problem-solving ability. Yet, despite this competence, performance varies greatly. In response, scholars and practitioners have sought ways to improve performance.

As a starting point, MTurk offers guidance to requesters such as "make instructions easy to read," "be as specific as possible in your instructions," and "include examples." In addition, researchers suggest using verification tasks [4], giving feedback to workers while they work [15], and breaking complex tasks into smaller simple subtasks, and distributing tasks to multiple workers in parallel [6]. With the rapid growth of workers and requesters on MTurk, researchers and practitioners are seeking new ways to improve performance while maintaining the desirable attributes of the platform (e.g., speed, low cost, and scalability). Yet, few researchers have considered an affective approach despite an empirical link between affect and problem-solving performance.

Affect

We use the term "affect" as a superordinate category that subsumes various affective phenomena, including emotions, moods, stress, and motivational states [16-18]. Affective stimuli, such as music, film, or images, may induce any number of affective states. Since it is not always possible to discriminate between affective states empirically, or even theoretically [20], we will use affect as an umbrella category to cover all states that reflect, at their core, changes in positive/negative valence and arousal [21].

Affect and Problem-solving

For decades, researchers have examined links between positive affect and problem-solving [8], [22], [23]. Scholars theorize that positive affect supports the broadening of attention to remotely connected ideas and facilitates the recombination of these ideas to solve problems requiring insight [24]. Isen [8] proposes that positive affect influences cognitive activity in three ways. First, positive

affect increases the number of cognitive elements available for processing, thereby increasing the number of elements available for association. Second, positive affect supports defocused attention, thereby broadening attention and increasing the number of cognitive elements considered relevant to the problem. Third, positive affect increases cognitive flexibility, thereby increasing the probability of association between diverse cognitive elements [8], [26].

In laboratory studies, Isen and colleagues have established an extensive repertoire (see [8] for a review) of ways to experimentally induce short-term positive affect in non-computing environments. In Isen's studies, participants are given gifts and treats, shown a clip of a comedy film, or played an excerpt from affect-laden music. For example, in two separate experiments with college students, participants were shown a clip of the movie *Gag Reel* or given a small bag of candy to induce affect [26]. In both cases, positive affect improved performance on insight problem-solving tasks, such as Mednick's Remote Associates Task (RAT) [27].

Affect and Motivation

Positive affect can also influence the cognitive processes underlying motivation [29]. Scholars posit that positive affect influences expectancy (belief that effort will result in good performance), which in turn influences goal commitment, goal setting, persistence, and performance [30]. In a study of 97 undergraduate students, researchers induced positive affect by giving the students a bag of candy and then asked them to solve 10 anagrams. The study found that participants induced with positive affect performed better, were more persistent, tried harder, and reported higher levels of motivation than those in the neutral condition who did not receive a bag of candy. It may therefore be possible to use positive affect to improve performance and motivation on crowdsourcing platforms, particularly those that use performance-contingent reward systems (such as MTurk).

Music and Affect

While affect can be elicited by many different kinds of stimuli, music is especially evocative. Music is often regarded as the "language of the emotions" [32] and studies report that the most common reason people listen to music is to influence their affective state [33].

Psychologists theorize that music evokes affect through a variety of different processes, including brain stem reflexes, evaluative conditioning, emotional contagion, visual imagery, episodic memory, and musical expectancy [33]. In the laboratory, music-evoked affect has been used to influence behavior [18], perception, [34], attention [34]. For instance, Rowe and colleagues used music to induce positive and negative affect in participants and found that participants who were exposed to happy music solved significantly more insight problems than those exposed to sad music [34]. In our work, we seek to understand whether music can induce affect in a real-world context – in this

case crowdsourcing – and if this influences performance. The research presented in this paper investigates the following possibilities: *Can affect be induced through music in an online crowdsourcing environment? If so, how might we induce positive affect in a contextualized way to improve task performance?*

Affect Computational Priming

Affect can be induced in a number of ways, ranging from lengthy (10min or more) mood induction procedures to microsecond subliminal presentations. In our work, we investigate affective priming – a technique that uses implicit mechanisms to induce changes in affect. Drawing from recent HCI research on affective priming in a digital context [37], we focus on two techniques: *active priming* and *passive priming*. We define active primes as primes that are attended to directly, even while their mechanisms and their influence remain implicit. Passive primes, by contrast, are not intended to capture attention; they exist in the background and not engaged with directly. Both methods involve supraliminal priming techniques, but the former is designed to engage attention, while the latter is not. To our knowledge, this distinction is not made explicit in the extant psychological literature, but we find it useful in delineating two important varieties of affective priming.

Based on existing literature linking positive affect with insight problem-solving performance, we formulated the following hypotheses:

Hypothesis 1

An *active* priming procedure with positive music will enhance problem-solving performance relative to a neutral condition with no primes.

Hypothesis 2

A *passive* priming procedure with positive music will enhance problem-solving performance relative to a neutral condition with no primes.

STUDY DESIGN

To study the influence of affect on task performance in micro-task crowdsourcing environments, we chose music as an affective prime, because music has been found to reliably influence affect in empirical laboratory studies [34].

Amazon Mechanical Turk

We selected MTurk as our micro-task platform because the platform is popular among researchers, practitioners, and workers [4]. While MTurk workers are often referred to as "turkers," when describing our empirical study, we refer to the workers as participants.

Stimuli

For music-based affective primes, we selected two instrumental pieces that have been used and validated in previous mood induction research [34], [38]. The positive prime was Bach's *Brandenburg Concerto No.3* (played by jazz flutist Hubert Laws) and the negative prime was

Prokofiev's *Alexander Nevsky: Russia Under the Mongolian Yoke* (performed by the Los Angeles Philharmonic and played at half-speed). In accordance with

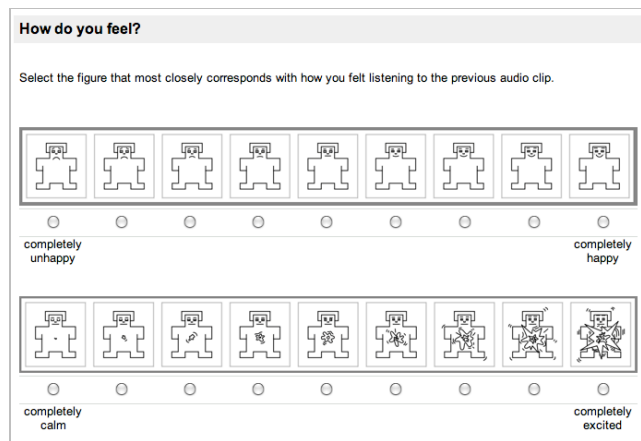


Figure 2. Screenshot of the Self-Assessment Manikin task on Mechanical Turk.

the views of other researchers in the field of music and affect, we do not believe a piece of music can be affectively neutral [18]. Therefore, we restricted our musical stimuli set to just positive and negative pieces.

We selected short excerpts from each piece that were particularly representative of the overall negative/positive valence of the piece. To address differences in average sound pressure levels between the two clips, peak signal values for both excerpts were normalized to -3dB.

To validate our music primes in a crowdsourcing context, we collected affect and arousal ratings for each excerpt on MTurk. 15 participants rated the positive musical excerpt and 15 participants rated the negative musical excerpt. We used the 9-point Self-Assessment Manikin (SAM) Scale, which uses graphical figures to depict human emotions ranging from smiling (i.e. happy) to frowning (i.e. unhappy). The SAM also depicts a range of arousal states, from completely calm to completely excited (see Figure 2). After listening to the 30-second audio clip on a web interface, participants were shown the SAM for emotion and arousal and were asked to “*Select the figure that most closely corresponds with how you felt while listening to the previous audio clip.*” This procedure targeted felt emotions actually experienced by the participants, rather than those merely perceived within the piece. The distinction between perceived and felt emotions is often blurred in research in music-evoked emotions, yet there is evidence that these types of responses can be differentiated empirically [39].

Our SAM results show that participants felt happier while listening to the positive music ($M = 7.8$, $SD = .86$) as compared to the negative music ($M = 4.47$, $SD = 1.19$); $t(28) = 8.80$, $p = .01$. Arousal ratings did not differ significantly between the two pieces.

EXPERIMENT 1: ACTIVE PRIME

Participants

186 participants from MTurk completed our first experiment. Because our task required mastery of the English language, we restricted enrollment to English speaking, U.S. workers with HIT qualifications of 95% or higher. 175 participants were included in our final analyses, after removing data from 21 non-U.S. participants,¹ participants who had completed earlier pilot versions of the study, and one participant who failed every test item (indicating poor task comprehension).

Measures

To measure performance, we created an online version of the Remote Associates Task (RAT) [40]. The RAT was used as a dependent measure because it has been developed and validated as a measure of insight-based problem-solving in hundreds of studies over the past 30 years [40].

In the RAT test, participants are shown three words and are asked to generate a fourth that forms a two-word phrase or compound with the other three. As an example, if given the words “aid,” “rubber,” and “wagon,” the subject would be expected to reply with the word “band”, as in “band-aid,” “rubber band,” or “bandwagon.” Reaching a solution requires insight, or what psychologists refer to as an “Aha” moment, a process that co-occurs with activation in the right hemisphere of the brain [28]. Psychologists find the RAT particularly useful for studying insight and creative performance, because it is thought to have high predictive validity – that is, achievement on the RAT is thought to correlate well with other creative problem-solving tasks [28].

While hundreds of RAT items have been published, we selected 20 of moderate difficulty, based on norms collected by previous researchers [40]. We purposely avoided extremely difficult/extremely easy RAT items, because we wanted to avoid the ceiling/floor effects described by Isen et al [26].

Procedure

We paid participants \$.35 to complete the task and our HIT description told potential workers that they would “*complete several short word puzzles.*” After accepting our HIT, participants read and agreed to an online consent form. Participants were shown how to complete the RAT task, and were given two examples with solutions provided.

After the instructions phase, participants were randomly assigned to one of three experimental conditions, each disguised as a verification task. In the neutral condition,

¹ We excluded individuals with IP addresses from outside the U.S. to help ensure that our participants were indeed English speakers.

participants were told: “To show us that you are paying attention, please write down today’s date.” In the positive

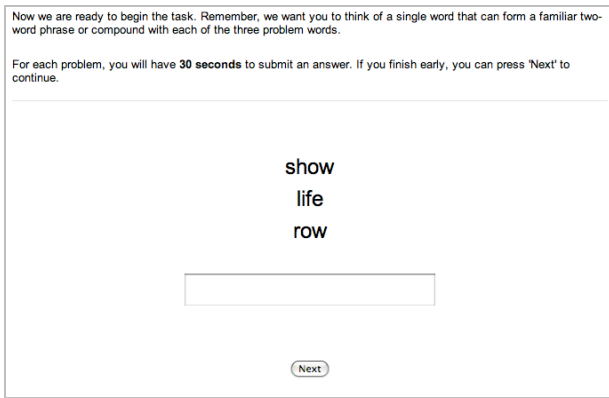


Figure 2. Screenshot of the RAT, as administered on MTurk.

and negative music-based priming conditions, participants were asked to listen to a 30-second audio clip and describe what they heard in the clip. Specifically, they were told: “To show us that you are paying attention, please turn up your computer volume and listen to the following 30 second clip. Please tell us: What do you hear in this audio clip?” Participants in the positive music condition heard an excerpt from Bach while those in the negative music-based prime condition heard an excerpt from the Prokofiev.

After completing the verification task, participants were reminded of the RAT procedure and were told that they would have 30 seconds to complete each word puzzle. They were also told that they could advance to the next question if they finished early (see figure 2). Our timer was controlled via javascript and was not visible to participants on-screen. The presentation order of the 20 RAT items was randomized for each subject.

After completing the 20 RAT items, participants were asked to report basic demographic information (including age, gender, and number of years speaking English, location).

Analysis

Table 1 shows the performance of our participants across each experimental condition. A one-way between subjects ANOVA was conducted to compare the effect of music-based affective primes on RAT task performance. The means reflect the average number of test items successfully completed out of 20 (see figure 3). In support of hypothesis 1, there was a significant effect of priming condition on RAT performance score [$F(2,173) = 3.95, p = .02$]. Post hoc comparisons revealed a significant difference between scores in the positive condition ($M = 13.51, SD = 4.41$) and scores in the neutral condition ($M = 11.47, SD = 4.57$), $p = .02$. Crowdworkers primed with happy music significantly outperformed those who were simply asked to write down the date. There was not a significant difference between the

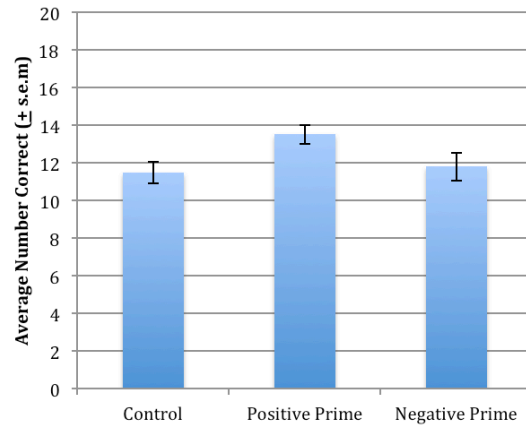


Figure 3. The average number of RAT items solved correctly (out of 20), per condition. The error bars reflect standard errors of the mean

Condition	% RAT Items Correct
Control	57.34
Positive Music	67.53
Negative Music	59.04

Table 1. The percentage of RAT items solved correctly, according to condition.

negative ($M = 11.81, SD = 3.89$) and positive conditions, $p = .11$, or the negative and neutral conditions, $p = .91$.

Discussion

Our results suggest that positive affect can prime performance on MTurk. Specifically, we find that participants primed with short excerpts of positive music outperform those given a neutral verification task with no primes. By simply asking workers to listen and attend to a short, 30-second piece of positive music, we were able to increase creative performance on the RAT by over 10%. By contrast, the negative music primes did not significantly affect performance.

In our experiment, MTurk workers performed similarly to participants in a controlled laboratory study. Using the same RAT items, Bowden & Jung’s [40] participants solved 70% correctly, on average. In our positive priming condition, in an uncontrolled crowdsourcing environment, our MTurk workers solved 67% of the RAT items correctly. This finding is surprising given the uncontrolled nature of MTurk. In a controlled laboratory study, participants meet face to face with the experimenter and thus may feel more obligated to perform well. Additionally, they are given

more instructions and example problems to learn from, prior to conducting the actual test.

Our results also show that affective priming techniques can be streamlined dramatically in order to suit the exigencies of online crowdsourcing environments. We improved performance using 30-second music excerpts, whereas previous research relied on much longer mood induction procedures. In the experiment conducted by Rowe et al. [34], participants listened to the positive and negative musical pieces for a full ten minutes, along with boosters throughout the experiment. Participants in that study were also instructed to generate affect-congruent thoughts while listening to the pieces. Our experiment demonstrates that even very short music pieces can act as affective primes and improve workers' performance. This has broad implications for HCI design and could also lead to new priming techniques in psychological research.

Crowdsourcing has challenged practitioners and researchers to reconsider what tasks are assigned and the time required to complete a task. According to best practices described by [4], MTurk tasks should be well bounded in terms of task content and task duration. As such, a verification task adds an extra step and may not be desirable for some crowdsourcing applications. With this in mind, we constructed another experiment and attempted to prime our participants throughout the task, without employing an a priori verification task. We refer to this as a passive priming technique [37], in the sense that workers are not actively attending to the prime, but are instead completing the task with the prime ever-present in the background.

EXPERIMENT 2: PASSIVE PRIME

Participants

Using the same exclusion procedure from experiment 1, we included 91 participants in our final analyses. Additionally, we removed data from subjects that had participated in the first study. Two other participants were also excluded because they did not answer any of the RAT items successfully and appeared to have misunderstood the instructions.

Measures

As before, we used the RAT task to assess insight-based problem-solving performance on MTurk.

Procedure

Our HIT paid \$.35 and was listed on MTurk with the same description from our previous experiment - "*solve several short word puzzles.*" After agreeing to our consent form, participants were introduced to the audio format of our HIT. They were told to wear headphones with the volume turned up to a moderate level, at which point they heard a neutral-sounding, female computer voice read aloud four different digits. Participants were asked to enter the digits into a textbox and they were not allowed to proceed until they had done so correctly. This procedure helped orient our

participants to the audio components of this HIT and it lent more credibility to our subsequent warning that some of the questions might be administered in an audio format (procedure described below).

Condition	% RAT Items Correct
Control	62.10
Positive Music	68.04
Negative Music	63.50

Table 2. The percentage of RAT items solved correctly in experiment 2, according to condition.

The participants were given the same RAT task instructions from experiment 1. Afterwards, they were shown a new screen and were told the following:

In this HIT, we are testing a new audio format. Most of the puzzles will appear on-screen, but additional questions may also be played in your headphones. It is therefore important to wear your headphones throughout the task. To test our audio stream, you may also hear some quiet hold music playing in the background.

These instructions served two important functions: (1) we needed to incentivize participants to listen to their computer audio throughout the entire task, and (2) in accordance with common priming methodologies [41], we wanted to hide the true purpose of the music stimuli and we therefore referred to it as "quiet hold music." In experiment 1, we were able to easily disguise the prime as a verification task. In both experiments, we asked participants to tell us about the purpose of the task and no one guessed that it was about the effect of music-evoked affect on performance.

Participants in the positive and negative music conditions heard full versions of the positive (Bach) and negative (Prokofiev) musical pieces. Since the RAT task could take as long as 10 minutes, we set our audio player to loop automatically, to fill the entire experimental session. Both pieces were mixed down, such that the peak signal value was -6dB below the audio captcha. Participants in the control condition did not hear any music or sounds playing in the background as they completed the task.

As soon as the task was over, four digits were read aloud and participants were told to write them down in a textbox provided on-screen. As before, the digits were read by a computer voice. The audio controls were hidden and our participants were not able to replay the clip. If they had the audio turned down or muted, they would not have been able to hear the digits. Thus, this *audio captcha* helped us discern whether or not participants were listening to the audio stream during the experiment. Participants who failed the audio captcha were automatically excluded from our dataset.

Analysis

Similar to experiment 1, we found that participants performed best when primed with positive music (Table 2).

To investigate this trend more closely, we used a one-way between subjects ANOVA to compare RAT performance across different background music conditions (positive music, negative music, no music). As with experiment 1, the mean for the positive music condition ($M = 13.61$, $SD = 3.02$) was greater than the means for the negative ($M = 12.7$, $SD = 4.40$) and control conditions ($M = 12.42$, $SD = 4.01$). However, the ANOVA for this experiment did not show a significant effect of priming condition on RAT performance [$F(2,88) = .835$, $p = .437$]. We therefore cannot confirm hypothesis 2: performance on the RAT did not change depending on the type of audio playing in the background throughout the task.

Discussion

While our results were in the direction we hypothesized, we did not observe a significant effect of primes on performance. In this experiment participants were not attending to the primes as deeply as they were in experiment 1, and this may account for the null effects. Contrary to experiment 1, participants were essentially told to ignore the music consider it as “quiet background music.” It is possible that the primes needed to be more salient to the worker. Future work could be done to adjust the volume of the primes, to see if performance changes as a function of audio amplitude.

Finally, we note that, as in experiment 1, our participants performed best when primed with positive music. While our results were not significant in experiment 2, it would be interesting to explore this trend further, perhaps with a larger sample size, with different musical stimuli, or by introducing the music during the task instructions and then continuing to play the music throughout the task.

General Discussion

To date, affective priming has typically been the purview of psychological research and has not been heavily explored in the context of HCI. Psychologists use affective priming techniques to help elucidate the workings of human affect, cognition, perception, and behavior. Here, we report significant affective priming effects in a real-world, HCI context. Specifically, we show that music-evoked affect can enhance performance on insight-based tasks conducted in an online crowdsourcing environment.

In our first experiment, using an active priming technique, we found that positive music improved performance in an online crowdsourcing environment. This finding aligns well with current research from psychology, which shows links between positive affect and performance. Our results extend findings from laboratory studies in psychology to an application domain that has relevance for the field of HCI. To our knowledge, we are also the first to show that a well-

validated insight-based problem-solving task, such as the RAT, can be administered successfully online on MTurk.

Experiment 1 showed significant performance gains using a simple 30-second verification task. By comparison, previous research on music priming and creative performance used a 10-minute induction procedure. Our findings thus show that priming techniques can be streamlined considerably for use on micro-task crowdsourcing platforms.

In our second experiment, we did not see a significant effect of priming condition on RAT performance. While there could be many reasons for these results, we suspect that quiet background music might be too subtle to induce a reasonable effect size. In our study, we found that an active priming method, deployed immediately prior to the focal task, achieved a more powerful effect.

Design Implications for Micro-Task Crowdsourcing Platforms

Our findings have implications for how designers implement crowdsourcing tasks. While much attention has been placed on the cognitive and motivational effects of crowdsourcing task designs, less attention has been placed on how affect might play a role. As our results show, affective features can influence performance, and should therefore be considered carefully. Moreover, our results indicate that positive affective primes could be leveraged to actually improve performance on insight-based tasks.

In the past, priming interventions were perhaps not terribly useful for most online work. But, as more and more people go online and contribute content, in a moment-to-moment, ad hoc fashion, priming techniques become quite relevant. While the temporal dynamics of primes are still poorly understood, they are not thought to linger for long periods of time [41]. But, on micro-task platforms, such as MTurk, many tasks require just a few short minutes of the worker’s time. Also, many insight-based crowdsourcing tasks, such as found in “Fold-it”, utilize a small, circumscribed problem space. The problems for which insight is required do not necessarily require long-term engagement. A short-term effect of a prime, aggregated over many hundreds, if not thousands of people, could therefore significantly improve the performance of the crowd.

Also, while experiment 1 used a verification task to embed our primes, there are may be other affect-delivery opportunities. For example, to reduce crowd latency, researchers are exploring ways to place workers on retainer, so that multiple workers can be ready to perform at the same time at any given instant [42]. As workers wait to complete the focal task, they are given an incidental task to keep them occupied and to ensure that they stay in the worker cue. The incidental task could serve as an affective prime and could ready workers to perform focal tasks at the highest possible level. For instance, workers on retainer could be asked to listen to positive music or humorous

videos before they are assigned to creative problem-solving tasks.

LIMITATIONS

In both experiments, we chose to focus on insight-based performance tasks – a type of task for which performance varies in crowdsourcing environments. While some crowdsourcing tasks require insight, many other tasks require analytic skills, or rote skills that simply tax patience and motivation. Future research should be conducted to see whether other crowdsourcing tasks could be positively influenced by affective priming techniques. We expect this to be the case, given findings from psychology that show effects of affective primes on many different performance domains, including attention [34], perception [43], and goal-directed behavior [44]. Still, we acknowledge that laboratory conditions are considerably different from real-world crowdsourcing environments and it is not clear whether the effects we observed with the RAT would hold up across many different performance domains. That said, we chose to use the RAT not for its ecological validity but instead for its predictive validity. Performance gains on the RAT task should extend to other insight-driven creative tasks. Conversely, performance gains on an idiosyncratic crowdsourcing task, without well-established norms, might fail to generalize to other types of tasks or other types of crowdsourcing platforms.

Our experiments are also limited because they did not explore other priming modalities beyond music (such as video, images, or text). It is therefore unclear whether or not other prime types could achieve similar results.

Additionally, our experiments were restricted to MTurk. We chose this platform because it has a wide user base, and because it can be a challenging setting to induce insight-based problem-solving work. However, because we restricted our experiments to MTurk, we do not know whether music-evoked affect might influence performance on other crowdsourcing platforms.

Finally, we must consider the ethics of priming in HCI. We argue that priming is ethical if inline with the workers' goals – in this case completing a task to a certain standard so that they can be paid. We cautiously warn against the use of priming that is out of line with workers' interests – such as priming concepts that might encourage workers to longer hours without appropriate compensation.

FUTURE RESEARCH

Our studies suggest a number of areas for future research. First, we plan to explore different affective priming modalities such as video, text, images and non-musical sound. Videos, for example, could be used to invite workers to complete tasks or non-musical sounds could be used during training. Videos may turn out to be stronger primes

because they stimulate both the visual and auditory senses. Other design features could also induce affect using colors, textures, and typography. For example, affect-laden icons could be used to direct workers through a task.

Second, we plan to investigate the role of affective primes in other labor markets. While Mechanical Turk is a popular crowdsourcing platform, several other online labor markets exist, such as oDesk and 99Designs. These markets offer different types of work, which may attract different types of workers willing to expend different amounts of cognitive and creative ability. Future work will consider if music also primes affect on these platforms and the extent to which mood induction procedures influence performance.

Beyond problem-solving, affect has been shown to influence the evaluation of others [45], negotiation skills [47], and helping behavior [48]. Consequently, affective priming may be useful peer-production platforms like Wikipedia that rely on people working effectively together.

In addition to affect, primes can be used to induce a variety of mental states that influence behavior. While this study focuses on priming affect, we can imagine computationally priming concepts that are congruent with a given technology's ideal behavior. The Amazon Kindle, for example, currently presents images of famous authors as its screen-saver. It is possible that this type of image primes general notions of literacy and literary achievement, and thus induces users to read more often and for longer time periods. However, to our knowledge, this possibility has yet to be tested empirically.

CONCLUSION

As HCI researchers, we must remember that affect, emotions, and moods (and indeed all affective phenomena) are a fundamental part of what makes us human. Since HCI is a humanistic field, it behooves us to consider the affective states of the user in relation to the technology of interest. Doing so will help us design better technologies, and it will open doors to many new research areas.

Crowdsourcing is an intriguing new area of research in HCI, and we believe it is particularly amenable to research on human affect. Importantly, affective components of crowdsourcing tasks may have a significant effect on performance. As HCI researchers explore this phenomenon in new ways, task design can be improved in ways that benefit both the crowdworker and the crowdsourcer.

REFERENCES

- [1] A. Quinn and B. Bederson, "Human Computation: A Survey and Taxonomy of a Growing Field," in *CHI 2011*, Vancouver, B. C., 2011.

- [2] S. Cooper et al., "Predicting protein structures with a multiplayer online game," *Nature*, vol. 466, no. 7307, pp. 756-760, 2010.
- [3] L. Yu and J. Nickerson, "Cooks or cobblers? crowd creativity through combination," in *CHI 2011*, Vancouver, B.C., 2011.
- [4] A. Kittur, E. Chi, and B. Suh, "Crowdsourcing User Studies With Mechanical Turk," in *CHI 2007: Proceedings of the ACM Conference on Human-factors in Computing Systems*, San Jose, CA, 2008.
- [5] W. Mason and D. J. Watts, "Financial incentives and the 'performance of crowds'," *ACM SIGKDD Explorations Newsletter*, vol. 11, p. 100-108, May. 2010.
- [6] M. S. Bernstein et al., "Soylent: a word processor with a crowd inside," in *Proceedings of the 23rd annual ACM symposium on User interface software and technology*, New York, NY, USA, 2010, p. 313-322.
- [7] S. Dow, A. Glassco, J. Kass, M. Schwarz, and S. Klemmer, "The Effect of Parallel Prototyping on Design Performance, Learning, and Self-Efficacy," in *TOCHI*, 2011.
- [8] A. Isen, "Positive Affect," in *The Handbook of Cognition and Emotion*, New York: Wiley, 1999, pp. 521-539.
- [9] J. Rogstadius, V. Kostakos, Aniket Kittur, B. Smus, J. Laredo, and M. Vukovic, "An Assessment of Intrinsic and Extrinsic Motivation in Crowdsourcing Markets," in *ICWSM11*, Barcelona, Spain, 2011.
- [10] E. L. Deci and R. M. Ryan, *Intrinsic Motivation and Self-determination in Human Behavior*. New York: Plenum, 1985.
- [11] L. von Ahn, "Games with a purpose," *Computer*, vol. 39, no. 6, pp. 92-94, 2006.
- [12] E. L. Deci, R. Koestner, and R. M. Ryan, "A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation," *Psychological Bulletin*, vol. 125, no. 6, pp. 627-668, 1999.
- [13] "mturk.com."
- [14] P. G. Ipeirotis, "Demographics of Mechanical Turk," *SSRN eLibrary*, Mar. 2010.
- [15] S. Dow, B. Bunge, T. Nguyen, S. Klemmer, A. Kulkarni, and B. Hartmann, "Shepharding the Crowd: Managing and Providing Feedback to Crowd Workers," in *CHI 2011*, Vancouver, B. C., 2011.
- [16] R. Picard, *Affective Computing*. Cambridge: MIT Press, 2000.
- [17] J. J. Gross and R. Thompson A., "Conceptual Foundations," in *Handbook of Emotion Regulation*, New York, NY, USA: Guilford Press, 2007, pp. 3-24.
- [18] P. N. Juslin and J. Sloboda, *Handbook of Music and Emotion: Theory, Research, Applications*, Reprint. Oxford University Press, USA, 2011.
- [19] N. H. Frijda, "The psychologists' point of view." Guilford, 2008.
- [20] L. F. Barrett, "Are Emotions Natural Kinds?," *Perspectives on Psychological Science*, vol. 1, no. 1, pp. 28 -58, Mar. 2006.
- [21] J. Russell, "Core affect and the psychological construction of emotion," *Psychological Review*, vol. 110, no. 1, pp. 145-172, 2003.
- [22] T. Amabile, S. Barsade, J. Mueller, and B. Staw, "Affect and Creativity at Work," *Administrative Science Quarterly*, vol. 50, pp. 367-403, 2005.
- [23] B. Frederickson, "What good are positive emotions?," *Review of General Psychology*, vol. 2, pp. 300-319, 1998.
- [24] B. Frederickson, "The broaden-and-build theory of positive emotions," *Philisophical Transactions of the Royal Society*, vol. 359, pp. 1367-1377, 2004.
- [25] A. Isen, "An Influence of Positive Affect on Decision Making in Complex Situations: Theoretical Issues with Practical Implications," *Journal of Consumer Psychology*, vol. 11, pp. 75-85, 2001.
- [26] A. Isen, K. Daubman, and G. P. Nowicki, "Positive affect facilitates creative problem-solving," *Journal of Personality and Social Psychology*, vol. 52, pp. 1122-1131, 1987.
- [27] M. T. Mednick, S. A. Mednick, and E. V. Mednick, "Incubation of creative performance and specific associative priming," *Journal of Abnormal and Social Psychology*, vol. 69, 1964.
- [28] E. M. Bowden and M. Jung-Beeman, "Aha! Insight experience correlates with solution activation in the right hemisphere," *Psychonomic Bulletin & Review*, vol. 10, 2003.
- [29] G. Clore, "Why emotions are felt.," in *The nature of emotion: fundamental questions*, New York: Oxford University Press, 1994, pp. 103-111.
- [30] A. Erez and A. Isen, "The Influence of Positive Affect on the Components of Expectancy Motivation," *Journal of Applied Psychology*, vol. 87, pp. 1055-1067, 2002.
- [31] R. Westermann, K. Spies, G. Stahl, and F. Hesse, "Relative effectiveness and validity of mood induction procedures: A meta-analysis," *European Journal of Social Psychology*, vol. 26, pp. 557-580, 1996.
- [32] D. Cooke, *The Language of Music*. Oxford University Press, USA, 1959.
- [33] P. N. Juslin and D. Västfjäll, "Emotional responses to music: the need to consider underlying mechanisms," *The Behavioral and Brain Sciences*, vol. 31, no. 5, pp. 559-575; discussion 575-621, Oct. 2008.
- [34] G. Rowe, J. B. Hirsh, and K. A. Anderson, "Positive affect increases the breadth of attentional selection," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, pp. 383-388, 2007.
- [35] B. Sollberger, R. Reber, and D. Eckstein, "Musical Chords as Affective Priming Context in a Word-

- Evaluation Task,” *Music Perception*, vol. 20, pp. 263-282, 2003.
- [36] T. Baumgartner, K. Lutz, C. F. Schmidt, and L. Jäncke, “The emotional power of music: how music enhances the feeling of affective pictures,” *Brain Research*, vol. 1075, no. 1, pp. 151-164, Feb. 2006.
- [37] S. Lewis, M. Dontcheva, and E. Gerber, “Affective computational priming and creativity,” in *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11*, Vancouver, BC, Canada, 2011, p. 735.
- [38] J. V. Wood, J. A. Salzberg, and L. A. Goldsamt, “Does affect induce self-focused attention,” *Journal of Personality and Social Psychology*, vol. 58, pp. 899-908, 1990.
- [39] M. Zentner and T. Eerola, “Self-report measures and models,” in *Handbook of Music and Emotion*, New York, NY, USA: Oxford University Press, 2010.
- [40] E. M. Bowden and M. Jung-Beeman, “Normative data for 144 compound remote associate problems,” *Behavior Research Methods, Instruments, & Computers*, vol. 35, pp. 634-639, 2003.
- [41] J. Bargh and T. L. Chartrand, “Studying the Mind in the Middle: A Practical Guide to Priming and Automaticity Research,” in *Handbook of research methods in social and personality psychology*, New York: Cambridge University Press, 2000, pp. 253-285.
- [42] M. S. Bernstein, J. Brandt, R. Miller, and D. R. Karger, “Crowds in Two Seconds: Enabling Realtime Crowd-Powered Interfaces,” in *24th Symposium on User Interface Software and Technology*, Santa Barbara, CA, 2011.
- [43] S. Murphy and R. B. Zajonc, “Affect, Cognition, and Awareness: Affective Priming with Optimal and Suboptimal Stimulus Exposures,” *Journal of Personality and Social Psychology*, vol. 64, pp. 723-739, 1993.
- [44] H. Aarts, R. Custers, and R. W. Holland, “The nonconscious cessation of goal pursuit: When goals and negative affect are coactivated,” *Journal of Personality and Social Psychology*, vol. 92, no. 2, pp. 165-178, 2007.
- [45] R. L. Cardy and G. H. Dobbins, “Affect and appraisal accuracy: Liking as an integral dimension in evaluating performance,” *Journal of Applied Psychology*, vol. 71, pp. 672-678, 1986.
- [46] A. Isen, T. Nygren, and G. Ashby, “The influence of positive affect on subjective utility of gains and losses: It’s not worth the risk,” *Journal of Personality and Social Psychology*, vol. 55, 1988.
- [47] P. Carnevale and A. Isen, “The influence of positive affect and visual access on the discovery of integrative solutions in bilateral negotiation,” *Organizational Behavior and Human Decision Processes*, vol. 37, pp. 1-13, 1986.
- [48] J. George and A. Brief, “Feeling good-doing good: A conceptual analysis of the mood at work - organizational spontaneity,” *Psychological Bulletin*, vol. 112, pp. 310-329, 1992.