Look, listen and learn: Exploring effects of passive entrainment on social judgements of observed others

Sarah Knight[1,2], Neta Spiro[1,3] and Ian Cross[1]

Abstract

Psychology of Music

Music is widely acknowledged to have social efficacy at the group level. This effect is hypothesised to be underpinned at least in part by entrainment. During collective musical behaviours, entrainment – the shared synchronisation of internal oscillators – is suggested to afford the perception of actions, intentions and motivational states as *joint* action, *shared* intentionality and *mutual* motivational states, which in turn fosters interpersonal affiliation and prosocial behaviours, including trust. However, it is unknown whether entrainment's effects on prosociality persist when we are passive observers. In this study, 44 participants (21 women; average age = 28; average years of musical training = 10) watched audio-visual tokens in which a) the footsteps of an actor were entrained (synchronised) with a drumbeat, b) the footsteps were disentrained (unsynchronised) with the drumbeat and c) the soundtrack was grey noise (control condition). Participants were subsequently required to decide if the actor was engaged in a trustworthy or untrustworthy activity. Results show that participants were more likely to judge the actor as trustworthy in the entrain condition than the disentrain condition, but that the entrain condition was not significantly different to the control condition. Furthermore, this pattern of results was only found for a subgroup of the stimuli. There were no effects of age, gender or musical training. Given the nature of the task, which encourages passive entrainment rather than active movement, these findings indicate that the prosocial outcomes of musical engagement may be more common and have a broader significance than previously suggested.

Keywords

audio-visual, beat perception, entrainment, social judgements, synchronisation

[1]Centre for Music and Science, Faculty of Music, University of Cambridge, UK [2]MRC Institute of Hearing Research, University of Nottingham, UK [3]Research Team, Nordoff Robbins, London, UK

Corresponding author:

Sarah Knight, MRC Institute of Hearing Research, University of Nottingham, University Park, NG7 2RD, UK. Email: sarah.knight.34@gmail.com

Music and prosociality

Music has frequently been postulated to have positive social effects at the group level – such as the promotion of interpersonal bonding, trust and altruism – and such suggestions have come from a variety of disciplines including neurobiology (e.g., Freeman, 2000), anthropology (Fiske, 2004) and even history (McNeill, 1995). One explanation of this phenomenon from the field of music psychology is that of Cross (2009). Cross highlights the importance in human musical experiences of what he terms *motivational-structural* principles – interpretations guided by the common links between organisms' motivational states and the acoustic characteristics of their signals – and *socio-intentional* processes – the interpretation and/or attribution of human agency and intentionality. However, Cross suggests that these alone are not enough to explain the prosocial and affiliative effects engendered by communal musical activity; to do this, one must look to entrainment.

Entrainment

Entrainment refers to the synchronisation of a biological, internal, regular oscillator with an external, periodically-recurring event. In two interdependent processes, it both allows expectancies to be generated about the timing of future events and also directs attention, matching attentional peaks to the external events (Barnes & Jones, 2000; Drake, Jones, & Baruch, 2000; Gjerdingen, 1993; Jones & Boltz, 1989; Large & Jones, 1999; London, 2012). Humans may make their neural entrainment manifest through entrained physical actions (Kirschner & Tomasello, 2009); for example, many people tap their feet to the regular "beat" when listening to a piece of music. However, neural entrainment needn't necessarily have physical correlates; it has been shown to be present in babies too young to actively synchronise their movements to a beat (Winkler, Háden, Ladinig, Sziller, & Honing, 2009) and experience suggests that it is perfectly possible to "keep up with" the beat in a piece of music without tapping one's foot. Nevertheless, even when there is no overt physical correlate, neural entrainment to a beat has been shown to activate motor regions in the brain, providing evidence for a strong perception/ action link (Grahn & Brett, 2007); this suggests a qualitative similarity between neural (passive) and physical (active) entrainment.

Entrainment, prosocial behavior and affiliation. During communal musical activities such as group singing, dancing and instrumental playing, the collective entrainment involved is suggested by Cross to increase the likelihood that the human action, intentionality and motivation derived through motivational-structural and socio-intentional interpretive processes will be perceived as *joint* action, *shared* intentionality and *mutual* motivations – and that this in turn fosters social bonds, interpersonal affiliation and prosocial behaviour (Cross & Woodruff, 2009). This view is shared by other authors. Bispham (2006) suggests that, through creating a shared temporal framework between participants in a group

activity, entrainment gives rise to collective emotionality, shared experience and group cohesiveness. Kirschner and Tomasello (2009) emphasise the social, cooperative role of entrainment and put its efficacy down to its ability to create joint attention between participants – the underlying condition necessary for shared psychological states and successful joint actions (Sebanz, Bekkering, & Knoblich, 2006; Tomasello & Carpenter, 2007). Comparable, though less grounded in cognitive theory or experiment, is McNeill's (1995) notion of "muscular bonding" – an evolutionarily beneficial enhancement of social cohesion caused by group movement in unison with a regular pulse. Hove and Risen (2009), meanwhile, treat the synchronous behaviour afforded by entrainment as related, but superior, to mimicry: where mimicry is shared action, synchrony is shared action in shared time. Thus, they argue, since mimicry fosters social bonding, affiliation and interpersonal closeness through the creation of shared self/action representations, synchrony (i.e., entrainment) should do the same, but to an even greater extent. Infant/caregiver entrainment, too, has been suggested to foster intersubjectivity, to give rise to shared emotions and experience and to regulate emotion beneficially (Cowley, Moodley, & Fiori-Cowley, 2004; Malloch, 1999–2000; Trehub, 2003; Trevarthen, 1999–2000). In short, temporal synchrony is seen as something special, going beyond "coordination" in a broader sense.

These theories have found confirmation in empirical studies: Hove and Risen (2009) discovered that synchronous tapping between participant and experimenter caused a significant increase in participants' affiliation ratings of the experimenter, compared to both a disentrainment condition – in which the tapping was asynchronous – and a control condition – in which the experimenter sat still; the disentrainment and control conditions were not significantly different from each other. Valdesolo, Ouyang, and DeSteno (2010) found increased feelings of "similarity" and "connectedness" between participants following a period of synchronised, as opposed to asynchronised, activity. Valdesolo and DeSteno (2011) subsequently confirmed this link between synchrony and judgements of similarity and liking, and also showed that synchrony led to higher levels of compassion and helping than asynchrony. Wiltermuth and Heath (2009) presented evidence that groups who had engaged in synchronised activity felt a stronger sense of group cohesion and trust than those who had not, and were also more likely to cooperate or behave altruistically in subsequent economic games. These findings have been extended to children aged 8–9 years by Rabinowitch and Knafo-Noam (2015), who found that pairs of children who engaged in synchronous rhythmic interaction perceived each other as more simi-lar and closer to themselves than children with whom they had taken part in similar but asynchronous interactions. Such prosocial effects appear to be present even in young children, with 4-year-olds showing a greater tendency towards spontaneous cooperative and helpful behaviour after engaging in entrainment-based games than after other communal activities (Kirschner & Tomasello, 2010). It is worth noting

that a recent follow-up cross-cultural study found no prosocial effects amongst 2–4 year-old children following a similar drumming game (Kirschner & Ilari, 2014). However, the authors suggest that this finding may be due to the limitations of their study design; and indeed other recent work using a different methodology found increased levels of prosocial behaviour following synchronous versus asynchronous bouncing for children as young as 14 months (Cirelli, Einarson, & Trainor, 2014). Finally, the work of Miles and colleagues has demonstrated a role for entrainment in facilitating social-cognitive processing and mediating affiliation. Macrae, Duffy, Miles, and Lawrence (2008) showed that when the physical gestures of a conversing dyad were synchronised to a regular external pulse participants' memory for both utterances and facial appearance was facilitated. Miles, Nind, and Macrae (2009) showed that synchronised, but not asynchronised, dyadic activity eliminated the normal memory advantage for self-relevant (as opposed to other-relevant) information, a finding recently substantiated in the real-world context of dance by Woolhouse, Tidhar, and Cross (2016), who employed silent disco technology (in which dancers wear headphones and dance alongside others who may be dancing to the same or to different soundtracks) to show that memory for person attributes is better for those dancing to the same track than for those dancing to different tracks. Miles, Griffiths, Richardson, and Macrae (2010) found that spontaneous synchrony was substantially reduced when the coordination partner was disliked by the participant; and a subsequent study (Miles, Lumsden, Richardson, & Macrae, 2011) showed that participants increased their synchrony with individuals from a notional "outgroup" when they anticipated having to engage in a dyadic social encounter with them later – they appeared to be trying to "smooth the path" ahead of the expected social encounter through synchrony.

Passive and covert scenarios

The theories cited above all describe the prosocial effects of entrainment as occurring via active, and interactive, entrainment processes. However, music is frequently experienced in situations in which one or more members of the interaction is a passive listener and/or observer – either throughout (as in the Western concert hall tradition) or temporarily. In these situations, entrainment, as an automatic neural response, will persist (Jones & Boltz, 1989); however, it is likely to involve no overt entrained behaviour at all, or at least a very limited and non-interactive set of covert physical responses (e.g., toe-tapping). It is unknown whether or not, in these constrained and non-interactive situations, the power of entrainment to influence affiliation and prosocial behaviour remains effective for listeners/observers.

On the one hand, the passive listener/observer is no longer engaged in overt physical action and/or interaction. Although they receive visual and audio

information from the human action(s) they are observing, they are not receiving any visual, audio, haptic or sensorimotor feedback indicating that they themselves are synchronised with these actions. And even if the listener *is* engaged in some covert form of entrained behaviour such as toe-tapping, this is an activity removed from social context: the entrained-to other is unaware of the listener's action and thus there is no possibility for negotiation or interaction (a symphony orchestra is extremely unlikely to change tempo because someone in the stalls is slowing down their foot tapping!). This poses questions: in the case of truly passive listening, through which medium would entrainment provide the sense of something "shared" – the crucial feeling of "togetherness" which seems to underlie its affiliative powers? After all, there can be no concept of "joint action" without action in the first place. And for the listener engaged in covert entrainment, would the evident lack of bidirectional awareness and consequent absence of interaction interfere with the creation of these perceptions? It seems plausible that for a true sense of "togetherness", actions and motivations need to be perceived not just as shared (i.e., the same) but in fact as mutual (i.e., reciprocal). After all, existing studies which provide evidence for the prosocial effects of entrainment do so through engaging participants in overt entrainment, usually in an interactive scenario and at the very least in a context in which relevant others are aware of participants' entrained behaviour.

On the other hand, one could suggest that, although there is no literal action, entrainment in the truly passive scenario could work instead towards a more abstract sense of "shared intentionality". That is, when the observer's neural entrainment is combined with the sight of a person apparently entrained to the same stimulus, a sense of shared goals and mutual "sameness" could arise, perhaps alongside a more complex sense of joint action removed from the actual physical sphere; these effects would in turn promote prosocial behaviour, social bonding and interpersonal affiliation. Indeed, given the strength and robustness of the social effects of active entrainment so far discovered, and the close perception/action ties observed in human beat perception (Grahn & Brett, 2007), such a suggestion does not seem implausible. In the case of covert entrainment, meanwhile, one could suggest that indications of synchrony with the observed other obtained through haptic and sensorimotor feedback are sufficient to induce in the listener a percept of shared action and intention, and that this percept – although not reciprocal – is in turn sufficient to generate prosocial outcomes.

To test these suggestions, levels of interpersonal trust were used as a means of measuring the prosocial effects of passive/covert entrainment. Interpersonal trust is an important part of group cohesion and strongly linked to prosocial behaviours (Eisenberg et al., 1989; Freeman, 2000; Rotenberg et al., 2005; Wentzel, 1991). Further to this, music's ability to elicit trust is proposed by Cross and his collaborators as one of the main processes underlying its positive social effects

(Hawkins, Cross, & Ogden, 2013).

Videos of a walking human figure were paired with one of the following soundtracks: a drumbeat that matched the figure's footfalls (entrained); the same drumbeat, but either too fast or too slow to match the figure's footfalls (disentrained); or grey noise (control). In each video, the figure was holding a different prop, and participants were required to decide whether or not that figure was involved in a trustworthy or untrustworthy activity. The primary hypothesis which this experiment seeks to test is that, during listening tasks which do not explicitly demand entrained behaviours and are removed from a social context, entrainment (as opposed to disentrainment) can nevertheless give rise to higher levels of interpersonal trust between the listener and those who can be observed entraining to the same stimulus.

Hypotheses

Main hypotheses

H1: Participants' responses will show a greater tendency towards judgements of trustworthiness in the entrained condition than in the disentrained (fast or slow) conditions.

All of the drumbeats used are rhythmically regular, so all allow for entrainment. However, in the entrained condition, the most salient beats are aligned with the footsteps of the videoed figure, giving the impression of the figure being entrained to the same stimulus as the listener. In the fast and slow conditions, meanwhile, none of the beats are aligned with the footsteps of the videoed figure, nor is there any other visible form of synchronisation between the two. Thus, although the figure still moves regularly, and the listener can still entrain to the drumbeat – ensuring similar levels of complexity across the three conditions – entrainment is not perceptibly shared between the listener and videoed figure in the fast and slow conditions.

As discussed above, links have been demonstrated between shared, active entrainment and prosocial behaviour, and prosocial behaviour has been shown to be linked to interpersonal trust; also, strong perception/action links have been observed in human beat perception, suggesting a close relationship between active and passive entrainment. It is therefore suggested that, during passive listening, higher levels of trust will emerge between the listener and the videoed figure when the figure is perceptibly entrained to the same stimulus as the listener (entrained condition) than when it is not (disentrained – fast and slow – conditions).

H2: Participants' responses may show a greater tendency towards judgements of trustworthiness in the entrained condition than in the noise condition.

Previous research (Hove & Risen, 2009) suggests that active, shared entrainment boosts positive interpersonal judgements relative to a control condition. The current study, by contrast, focuses on passive and/or covert entrainment, removed from a social context. The implications of this for the relationship between the entrained and control conditions is unclear. Nevertheless, as a working hypothesis, it is suggested that the entrained condition will produce higher levels of trust than the control condition.

Effects of tempo on trustworthiness judgements

Although the fast and slow conditions are predicted to produce more "untrustworthy" responses overall than the entrained condition, there may also be response differences between the fast and slow conditions themselves. The faster tempo may give listeners an impression of haste and/or urgency, which may imply nervousness or guilt; this may in turn lead to more "untrustworthy" responses in the fast condition than in the slow condition. However, this suggestion is tentative, and is not a main hypothesis.

Method

Materials

Eleven different videos were made of an actress walking whilst holding one of a variety of props, thus forming 11 different "scenarios".[1] Each scenario was intended to enable the actress's behaviour to be interpretable as either positive (trustworthy) or negative (untrustworthy). The actress was filmed from the same angle in every scenario, and both her clothing – black trousers and a black jumper – and also the surroundings – a green, urban space – were kept constant. Using a metronome as a guide, the actress walked at a regular tempo of crotchet = 116 – a tempo judged to be both comfortable and natural, and also neutral with respect to intention. After the videos were taken, the actress's face was blurred out using Final Cut Pro (see Figure 1).

Using Logic Pro, four different soundtracks were developed and matched with the videos to create four conditions (see Table 1). Three of the four conditions used a drumbeat which featured a low drum, a high bell and an untuned scraper, presented in the order bass-untuned- treble-untuned so as to create the sense of a subdivided 2/4 metre (see Figure 2). The fourth condition used grey noise; this condition was designed as a control that included no rhythmic information, but which nevertheless still contained an audio stimulus.

Figure 1. Still image from one of the videos used. In each scenario the background and figure were kept constant, while the prop (in this case a picture) was varied.

Table 1. Summary of the four conditions.

Condition	Description
1. Entrained	Drumbeat presented at 116 bpm, so that the bass and treble beats coincided with the actress' footfalls. As an introduction, the video was preceded by two full bars of the drumbeat.
2. Fast (disentrained)	Drumbeat presented at 145 bpm. The first bass beat coincided with the actress' first footfall, but there were no further coincidences between rhythm and footsteps. As an introduction, the video was preceded by two full bars of the drumbeat.
3. Slow (disentrained)	Drumbeat presented at 87 bpm. The first bass beat coincided with the actress' first footfall, but there were no further coincidences. As an introduction, the video was preceded by two full bars of the drumbeat. Grey noise.

4. Noise	As an introduction, the video was preceded by grey noise with the same duration as the introduction in the Entrained condition.

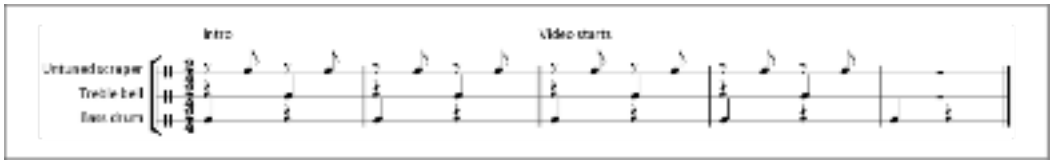

Figure 2. The rhythmic pattern used in conditions 1–3.

The video clips themselves had a relatively short duration (c. 2.5 s); all conditions therefore contained an audio introduction before the video started, during which participants saw a black screen. This ensured that all participants had sufficient exposure to the drumbeat in the first three conditions to give them the chance to entrain. In the fourth condition, an audio introduction of grey noise was used.

Every scenario (11 in total) was matched with every soundtrack (4 in total), creating 44 complete stimuli, called here "tokens".

The 44 tokens were divided up into four presentation lists such that:

- every list contained 11 tokens
- no token was repeated across lists
- no list contained the same scenario (i.e., visual scenario, not audio soundtrack) twice
- each list contained at least two tokens in each condition Each participant was assigned pseudo-randomly to one of the four lists such that equal numbers of participants were assigned to each list. The tokens within the assigned list were always presented in a fully random order. *Participants* The participants were 44 volunteers (average age = 28, *SD* = 15; 21 women; average years of musical training = 10, *SD* = 6)[2] recruited from the University of Cambridge and surrounding

Table 2. Props used in each of the 11 scenarios and the matched forced-choice two-option questions presented to participants.

Prop held	**Question asked: "Is/has she..."**	
(i.e., scenario)	*Trustworthy option*	*Untrustworthy option*
1. Bin bag	throwing out the rubbish?	fly-tipping in the countryside?

2. Briefcase	heading back home from work?	going to collect blackmail money?
3. Drainpipe (section)	going to help her friend with DIY?	planning to use it as a weapon?
4. Bunch of flowers	taking them to her elderly grandmother?	taking them to her illicit lover?
5. Football	bringing it to her local football team so they can play?	confiscated it from some kids who were just having some fun?
6. Bags (hold-all and backpack)	leaving to go to university?	been thrown out for stealing?
7. Picture (painting)	bringing it over for a friend?	just stolen it from a house?
8. Saw	about to help her friend prune the hedge?	about to steal a bike?
9. Shopping bag	carrying her shopping back to the house?	just snatched it from an old lady?
10. Spade	going to do some gardening?	about to bury a corpse?
11. Bag of sweets	bringing them to a children's party?	taken them away from a child?

Note. The prompt sentence ("You'll see a video of someone carrying [a prop]") used the name of the prop as given in this table; bracketed information is provided here for clarity only.

area via email advertisements and word of mouth. A between-subjects ANOVA revealed no significant differences between the four groups of participants assigned to the different presentation lists in terms of age, $F_{3, 37} = 1.5$; $p > .1$ or years of musical training, $F_{3, 37} = 0.31$; $p > .1$. A chi-squared test revealed that the number of females and males was not significantly different between groups, $\chi^2(3, N = 41) = 1.29$, $p = .73$. All participants provided informed consent and all procedures were performed in compliance with relevant laws and institutional guidelines. The experiment was given ethical approval by the Cambridge Psychology Research Ethics Committee.

Procedure

The experimental materials were presented via a laptop computer and on-ear headphones. Before viewing each token, participants were prompted with the phrase "You'll see a video of someone carrying [a prop]", with the name of the relevant prop (as per Table 2) inserted. Having viewed each token, participants were asked to make a judgement related to the trustworthiness of the figure in the video via a forced-choice two-option task. For example, they were asked to decide whether the person walking whilst holding a painting a) had stolen it (the "untrustworthy" option) or b) was bringing it over for a friend (the "trustworthy" option). A full list of the questions presented to participants, matched to the

relevant props, is provided in Table 2. The order of presentation of the "trustworthy" and "untrustworthy" answer options was varied randomly across all trials. After completing the experiment, participants were asked what they thought the purpose of the experiment was; having provided an answer, they were fully debriefed about its true purpose.

Results

General findings

For the purposes of analysis, "trustworthy" responses were assigned a value of 1 and "untrustworthy" responses a value of 0. All analyses were carried out using IBM SPSS Statistics 21. Each participant had been randomly assigned to one of four presentation lists (see Materials). A binary logistic regression which included each participant's mode response as a case and the four lists as categorical predictors revealed no significant differences between lists, so they were collapsed for the purposes of analysis.

As Figure 3 shows, the entrained condition produced higher overall trustworthiness ratings than either the fast condition or the slow condition, and marginally higher overall trustworthiness ratings than the noise condition. To test the significance of these differences, a generalized linear mixed model (GLMM) with a binomial distribution and logit link was run. This model included condition, age, gender and years of musical training as categorical fixed effects[3], and participants as random intercepts. No significant effects of the three demographic factors were found, so they were removed from the model. The subsequent model therefore contained only condition as a fixed effect and participants as random intercepts.

Overall, there was a significant main effect of condition, $F_3 = 2.938$, $p = .03$. Furthermore, significance testing of the log odds ratios (fixed coefficients) revealed significantly higher overall trustworthiness ratings for the entrained condition than either the fast condition or the slow condition, $t(480) = 2.527$; $p = .01$, odds ratio = 1.99 and $t(480) = 2.069$; $p = .04$, odds ratio = 1.76 respectively, thus supporting the main hypothesis advanced here (H1).

No significant difference was found between responses to the entrained condition and responses to the noise condition, $t(480) = 0.519$, $p > .5$. This does not support H2. Finally, no significant difference was found between responses to the fast condition and responses to the slow condition, $t(480) = –0.466$, $p > .5$.

Effect of scenario

When each scenario was examined individually, it became apparent that this overall response pattern was not found for each one: some scenarios clearly mirrored the overall

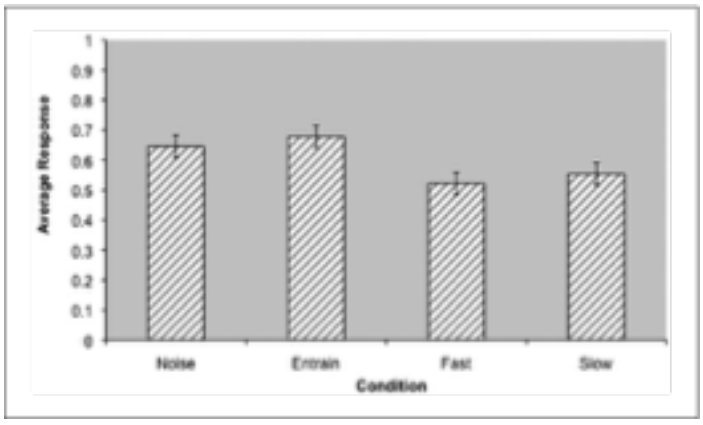

Figure 3. Average trustworthiness judgements for all stimuli. Error bars represent standard error here and throughout.

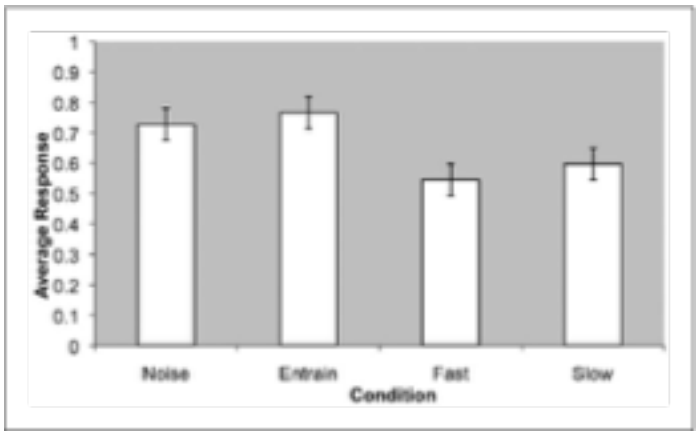

Figure 4. Average trustworthiness judgements for confirmatory stimuli.

pattern described above, while others did not. The scenarios could therefore be divided into two groups: confirmatory (i.e., confirmed the pattern) and non-confirmatory. The confirmatory group contained 7 scenarios (scenarios 1, 2, 3, 7, 8, 9 and 10; see Table 2) and the nonconfirmatory group the remaining 4 scenarios (4, 5, 6 and 11). Based on this visual analysis, the data was divided into these two groups, and the GLMM was re-run using the data from each of these groups of scenarios in turn.[4] Although the confirmatory and non-confirmatory groups were of different sizes, the non-confirmatory group was still considered to have a sufficient number of data points given the simplicity of the model (only one fixed and one random effect). Age, gender and years of musical training were initially included in both re-runs, but no significant effects were found, so they were removed from subsequent models.

Re-running the GLMM using only those scenarios which confirmed the original pattern produced – unsurprisingly – the same pattern of significant differences as the complete data set (see Figure 4). There was a main effect of condition, $F_3 = 3.589$, $p = .01$, and the differences between responses to the entrained and fast conditions, and between responses to the entrained and slow conditions, remained significant, $t(304) = 2.852$, $p = .005$, odds ratio = 2.77 and $t(304) = 2.146$, $p = .03$, odds ratio = 2.23. No significant difference was found between responses to the fast and slow conditions, $t(304) = -0.623$, $p > .5$; and no significant difference was found between responses to the entrained and noise conditions, $t(304) = 0.609$, $p > .5$.

However, as Figure 5 illustrates, the responses to the non-confirmatory scenarios showed no effect of condition: re-running the GLMM using only the non-confirmatory scenarios removed the main effect of condition, $F_3 = 0.07$, $p > .5$, and no further comparisons were significant (all $p > .5$).

Discussion

The finding that the entrained condition yielded significantly higher levels of trust than either of the disentrained conditions suggests that participants were more likely to trust, or associate trustworthy behaviour with, the videoed figure when that figure was moving to the same beat that the participant was hearing. This provides support for the main hypothesis – that, during

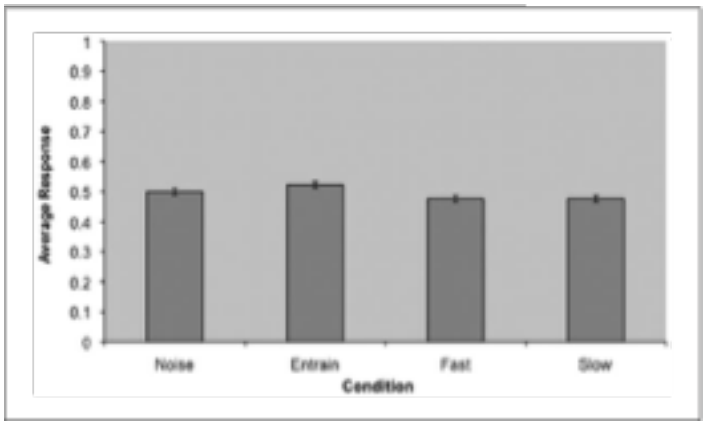

Figure 5. Average trustworthiness judgements for non-confirmatory stimuli.

passive listening, participants' responses would show a greater tendency towards judgements of trustworthiness in the entrained condition than in the fast or slow conditions. However, no significant difference was found between the entrained and noise conditions; thus the second hypothesis – that entrainment would boost trustworthiness judgements above the control condition – is unsupported. In fact it suggests that, in this case, entrainment is not promoting trust; rather,

disentrainment is having a negative effect on trust levels. This runs counter to the result of Hove and Risen's study, in which entrainment raised the level of prosocial outcomes above a control condition while disentrainment was not significantly different to the control (Hove & Risen, 2009). One possible explanation for this divergence is the differing nature of the control conditions in the two experiments: the current study used grey noise, whereas Hove and Risen asked the experimenter to remain still and create no auditory stimulus at all. In the latter case, participants might have felt that the experimenter was behaving in an unfriendly or aloof manner, and thus felt disaffiliated. In other words, this control condition may not have provided a true baseline measure, since it may itself have had socially negative consequences – an outcome which seems unlikely to have arisen from the grey noise condition used here. However, the disparity may also be because Hove and Risen focused on active rather than passive entrainment. Further research incorporating active entrainment into the current paradigm is needed to test these possibilities. Until the nature of the control conditions is fully understood, it will not be clear whether entrainment promotes trust relative to these control conditions or whether disentrainment lessens trust.

The absence of a significant difference between the fast and slow conditions undermines the tentative suggestion that the fast tempo might suggest some sense of urgency or a physiological response such as a racing heartbeat, both of which could imply guilt. Rather, this finding adds strength to the support for the main hypothesis, since it suggests that the pattern of results observed is due to the relative effects of entrainment versus disentrainment rather than to any effect of tempo.

The differing response patterns found for the confirmatory versus non-confirmatory scenarios are more difficult to interpret. However, it is suggested that this difference reflects the nature of the scenarios themselves: those scenarios that confirm the overall response pattern could be considered rather improbable, whereas the non-confirmatory scenarios are somewhat more probable. For example, the "painting" video fell into the "improbable" scenario type, since one is unlikely to see a thief walking openly with the stolen object, whereas the "football" video fell into the "more probable" scenario type, since the "untrustworthy" option was merely confiscating a football; similarly, the "briefcase" scenario was considered "improbable", since most people carrying briefcases are not blackmailers, whereas the "bag of sweets" scenario was more probable, since one might well have confiscated them from a child (see Table 2 above).[5] This difference in probability could cause participants to engage in qualitatively different types of responding. For example, it has been shown that social judgements of observed others are more likely to be influenced by low-level, peripheral information when personal relevance to the observer is low (Petty & Cacioppo, 1979) or when the observer is physiologically aroused (Sanbonmatsu

& Kardes, 1988). In this case, the untrustworthy options in the "improbable" scenarios are both somewhat surprising and also highly unlikely to have direct relevance to participants' own experiences, and therefore might prompt responses which are more susceptible to the effects of peripheral attendant phenomena like the drumbeats. The more commonplace "more probable" scenarios, on the other hand, seem likely to prompt responses which are more considered and more liable to draw on participants' own experiences; and these more intellectualised responses are less likely to be affected by peripheral attendant phenomena like the drumbeats. However, this interpretation is speculative, and further investigation would be needed to draw firm conclusions.

No participant identified the true purpose of the experiment. This speaks against these results arising from demand characteristics. In particular, most participants noticed that some drumbeats were faster or slower than others, and suggested that the study was investigating effects of tempo, yet there was no significant difference between the fast and slow conditions. Only one participant explicitly mentioned the difference between "in time" and "out of time" soundtracks, but again thought that the main purpose of the experiment was tempo-related. Some participants commented that the soundtrack, although audible, had not impinged on their thoughts at any point during the experiment.

Underlying mechanisms

It is hypothesised here that trust arises from a match between one's own entrainment processes and those of another. This match is suggested to give rise to a particular kind of shared intentionality which fosters social bonds, interpersonal affiliation and prosocial behaviour. In the case of passive musical engagement, this interaction is suggested to be "virtual": despite the absence of actual action and interaction, the shared entrainment processes nevertheless allow the observer to develop a sense of shared goals and action plans with regard to the observed other, which in turn promotes trust.

This experiment provides some initial support for the modulation of trust levels during passive engagement, although the direction of the effect (entrainment as beneficial versus disentrainment as detrimental) is unclear; furthermore, the study was not designed to test underlying mechanisms. It is therefore not known what caused the observed pattern of results, and whether or not the "virtual interaction" mechanism proposed here could be involved. One possible explanation for the observed results relates to processing fluency. As described above, entrainment to an auditory stimulus serves to direct attention to particular points in time, allowing for maximally efficient processing of events occurring at those points (Jones & Boltz, 1989). It has been demonstrated that this attention-driven facilitation of processing occurs cross-modally, with metrically-structured auditory stimuli

reducing response times to visual images when their presentation coincides with salient points in the metrical structure (Escoffier, Sheng, & Shermer, 2010). Increased processing fluency has been shown to give rise to more positive attitudes towards the processed object (Lee & Labroo, 2004; Whittlesea, 1993; Zajonc, 1968). Thus it could be the case that the actor in the disentrained conditions was judged less positively than the actor in the entrained and control conditions simply because the tokens in the disentrained conditions were processed less fluently due to the mismatch between the visual image and the entrainment induced by the auditory stimulus. In the other conditions, such a mismatch did not (entrained) or could not (control) occur. Interestingly, recent work suggests that it is synchrony per se that is key in fostering prosocial attitudes rather than the occurrence of action at temporally predictable (and hence high-attention) moments in time: Cirelli et al. (2014) compared the effects of bouncing babies in an asynchronous versus synchronous fashion using both temporally predictable and unpredictable stimuli, and found increased prosociality for both synchrony conditions, regardless of temporal predictability. Nevertheless, the authors note that the regular beat of music provides a particularly effective context for facilitating synchrony. Indeed, it may be the case that, amongst adult listeners controlling their own movements, a combination of temporal predictability and processing fluency during musical activities serves to promote particularly robust prosociality above and beyond that arising from synchrony alone. This is, however, speculative, and requires further testing.

Another possible explanation for the results observed here relates to affect. It could be suggested that the audio-visual mismatch in the two disentrained conditions prompted irritation or other negative affective responses amongst participants, thus creating a generalised tendency towards negative judgements that was not specifically related to the observed figure. Such an explanation would account for the unpredicted finding that disentrainment lowered trustworthiness judgements relative to a non-rhythmic control condition. Mismatch between the audio and visual components of a film is indeed reported as being irritating for viewers (CCIR, 1990); however, it seems unlikely that such irritation arose here or caused the results observed. Irritation is usually reported for a temporal discrepancy of a constant duration between a depicted activity and the sound it generates, such as asynchrony between a speaker's face and voice (e.g., Summerfield, 1992) or between a sequence of events and the associated soundtrack (e.g., Žonja, Livun, & Jambrošić, 2006). In other words, negative affect appears to arise as the result of an artificial temporal lag between two components which viewers normally perceive not just simultaneously, but as causally related – and, moreover, as two facets of a single multi-modal event. In the case of the disentrainment conditions used here, the temporal discrepancy between the footfalls and the accented beats was not consistent, but fluctuated considerably over the course of each video; moreover, it was quite clear that the

sounds – which included not only drumbeats but also bells and off-beat scrapers – were not being generated by the movements of the videoed figure. It therefore seems highly unlikely that participants would have been trying to interpret the audio and visual events as causally linked or to process them as facets of the same perceptual event; as a result, it also seems unlikely that the asynchrony would have proved irritating. Furthermore, no participant reported any annoyance or irritation; indeed, several commented that the soundtrack had not impinged on their thoughts at all (see above). Nevertheless, such a hypothesis needs to be tested before it can be entirely disregarded. On the other hand, there is evidence that increased processing fluency is associated not only with positive attitudes but also with a generalised positive affective response (e.g., Schwarz, 2004; Winkielman & Cacioppo, 2001; Winkielman, Schwarz, & Nowak, 2002), and there is some evidence that this kind of positive affective response can lead to the emergence of prosocial behaviour (North, Tarrant, & Hargreaves, 2004). If the entrained condition, but not the other conditions, was found to induce mildly positive affective responses in participants – presumably due to increased processing fluency – then this could also explain the results obtained here to some extent. Either way, the possibility that an affective response – positive or negative underlies these findings needs to be further investigated before firm conclusions can be drawn regarding underlying mechanisms.

Conclusion

The results of this experiment provide some initial evidence to support the suggestion that passive and/or covert entrainment can serve to modulate trust on the part of the listener/observer towards an individual who can be observed actively entraining to the same auditory stimulus as the listener. Given the passive nature of the task, these results suggest that the prosocial outcomes of musical engagement indicated by earlier studies may have a more widespread occurrence and significance than previously realised.

Acknowledgements

We are grateful to Matt Reed for writing the experimental software to Jonathan Green and Michelle Phillips for technical assistance, to the members of the Centre for Music and Science at the University of Cambridge, and to all the volunteers who took part in the experiment. Parts of this article have previously appeared in Knight, Spiro, Cross, and Reed (2011).

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported in part by a Domestic Research Studentship awarded by Selwyn College, Cambridge and the

University of Cambridge.

Notes

1. The research materials used in this study can be obtained by contacting the corresponding author.

2. Three participants did not complete the questionnaire so their age, gender and musical experience are unknown.

3. Age was divided into the following categories: 20 or under; 21–30; 31–40; 41–50; over 50. Years of musical training was divided into the following categories: less than 5 years; 5–9 years; 10 or more years. This was done to enable more straightforward post hoc testing of these variables and their interactions in the GLMM if necessary. In practice, the same overall significance patterns were obtained regardless of whether the continuous or categorical versions were used.

4. An alternative procedure is to run the model using all the data and incorporating scenario (confirmatory vs. non-confirmatory) as a second fixed effect alongside condition. Such an approach treats scenario type as a planned contrast, whereas it was not in fact part of the original study design and the decision to examine it was entirely post hoc. Nevertheless, since this approach allows for a true comparison between scenario types it was used to test whether there was indeed a difference between the confirmatory and non-confirmatory scenarios. Demographic factors were initially included as fixed effects, but removed when no significant main effects were found. The model produced a main effect of scenario type, indicating that trustworthiness ratings were significantly higher overall for the confirmatory than non-confirmatory scenarios, $F_1 = 13.774$, $p < .001$. There was no significant main effect of condition in this model, and no significant condition x scenario interaction, $F_3 = 2.187$, $p > .05$ and $F_3 = 0.880$, $p > .1$ respectively. Nevertheless, significance testing of the log odds ratios revealed that the significant differences between conditions reported for the complete data set were also present for the confirmatory scenarios: significant differences were found between responses to the entrained and fast conditions, and between responses to the entrained and slow conditions, $t(476) = 2.903$, $p = .004$, odds ratio = 2.85 and $t(476) = 2.146$, $p = .32$, odds ratio = 2.27, while no significant difference was found between responses to the fast and slow conditions or the entrained and noise conditions, $t(476) = –0.636$, $p > .5$ and $t(476) = 0.619$, $p > .5$). For the non-confirmatory scenarios, however, none of these comparisons reached significance (all $p > .5$). This corresponds to the findings for the two separate models reported in the main text. Taken together, these results suggest that, although the condition x scenario interaction did not reach significance, there were nevertheless somewhat different response patterns across conditions in the different scenario types.

5. This suggestion is further supported by the finding that trustworthiness ratings were significantly higher overall for the confirmatory as opposed to non-confirmatory

scenarios (see Note 3). This implies that participants were generally less likely to choose the untrustworthy option in the confirmatory scenarios, which in turn suggests that in these scenarios this option was relatively improbable.

References

Barnes, R., & Jones, M. R. (2000). Expectancy, attention, and time. *Cognitive Psychology*, *41*, 254–311. Bispham, J. (2006). Rhythm in music: What is it? Who has it? And why? *Music Perception*, *24*(2), 125–134.

CCIR. (1990). Tolerances for transmission time differences between the vision and sound components of a television signal. In *ITU-T Recommendation J.100 (06/90)*. Düsseldorf, Germany

Cirelli, L. K., Einarson, K. M., & Trainor, L. J. (2014). Interpersonal synchrony increases prosocial behavior in infants. *Developmental Science*, *17*(6), 1003–1011.

Cowley, S. J., Moodley, S., & Fiori-Cowley, A. (2004). Grounding signs of culture: Primary intersubjectivity in social semiosis. *Mind, Culture and Activity*, *11*(2), 109–132.

Cross, I. (2009). The evolutionary nature of musical meaning. *Musicae Scientiae*, *13*(2 Suppl.), 179–200. doi:10.1177/1029864909013002091

Cross, I., & Woodruff, G. E. (2009). Music as a communicative medium. In R. Botha & C. Knight (Eds.), *The prehistory of language* (Vol. 1, pp. 113–144). Oxford, UK: Oxford University Press.

Drake, C., Jones, M. R., & Baruch, C. (2000). The development of rhythmic attending in auditory sequences: Attunement, referent period, focal attending. *Cognition*, *77*, 251–288.

Eisenberg, N., Fabes, R. A., Miller, P. A., Fultz, J., Mathy, R. M., Shell, R., & Reno, R. R. (1989). The relations of sympathy and personal distress to prosocial behaviour: A multimethod study. *Journal of Personality and Social Psychology*, *57*, 55–66.

Escoffier, N., Sheng, D., & Schirmer, A. (2010). Unattended musical beats enhance visual processing. *Acta Psychologica*, *135*, 12–16.

Fiske, A. P. (2004). Four modes of constituting relationships: Consubstantial assimilation; space, magnitude, time, and force; concrete procedures; abstract symbolism. In N. Haslam (Ed.), *Relational models theory: A contemporary overview* (pp. 61–146). Mahwah, NJ: Erlbaum.Freeman, W. J. (2000). A neurobiological role of music in social bonding. In N. L Wallin, B. Merker, & S. Brown (Eds.), *The origins of music* (pp. 411–424). Cambridge, MA: MIT Press.

Gjerdingen, R. O. (1993). “Smooth” rhythms as probes of entrainment. *Music*

Perception, *10*(4), 503–508. Grahn, J. A., & Brett, M. (2007). Rhythm and beat perception in motor areas of the brain. *Journal of Cognitive Neuroscience*, *19*(5), 893–906.

Hawkins, S., Cross, I., & Ogden, R. (2013). Communicative interaction in spontaneous music and speech.

In M. Orwin, C. Howes, & R. Kempson (Eds.), *Music, language and interaction* (pp. 285–329). London, UK: College Publications.

Hove, M. J., & Risen, J. L. (2009). It's all in the timing: Interpersonal synchrony increases affiliation. *Social Cognition*, *27*(6), 949–961.Jones, M. R., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review*, *96*(3),

459–491.Kirschner, S., & Ilari, B. (2014). Joint drumming in Brazilian and German preschool children: Cultural

differences in rhythmic entrainment, but no prosocial effects. *Journal of Cross-Cultural Psychology*, *45*(1), 137–166. doi:10.1177/0022022113493139

Kirschner, S., & Tomasello, M. (2009). Joint drumming: Social context facilitates synchronization in preschool children. *Journal of Experimental Child Psychology*, *102*, 299–314.

Kirschner, S., & Tomasello, M. (2010). Joint music making promotes prosocial behavior in 4-year-old children. *Evolution and Human Behavior*, *31*(5), 354–364.

Knight, S., Spiro, N., Cross, I., & Reed, M. (2011). Exploring prosocial effects of entrainment in passive listening. In A. P. Ghiena, P. Jacquier, M. Valles, & M. Martínez (Eds.), *Musicalidad Humana: Debates actuales en evolución, desarrollo y cognición e implicancias socio-culturales. Actas del X Encuentro de Ciencias Cognitivas de la Música*, 469–476.

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How we track time-varying events. *Psychological Review*, *106*(1), 119–159.

Lee, A. Y., & Labroo, A. (2004). Effects of conceptual and perceptual fluency on affective judgment. *Journal of Marketing Research*, *41*, 151–165.

London, J. (2012). *Hearing in time: Psychological aspects of musical metre* (2nd ed.). Oxford, UK: Oxford University Press.

Macrae, C. N., Duffy, O. K., Miles, L. K., & Lawrence, J. (2008). A case of hand waving: Action synchrony and person perception. *Cognition*, *109*, 152–156.

Malloch, S. N. (1999–2000). Mothers and infants and communicative musicality. *Musicae Scientiae, 3*(1 Suppl.), 29–57. doi:10.1177/10298649000030S104

McNeill, W. H. (1995). *Keeping together in time: Dance and drill in human history*. Cambridge, MA: Harvard University Press.

Miles, L. K., Griffiths, J. L., Richardson, M. J., & Macrae, C. N. (2010). Too late to coordinate: Contextual influences on behavioral synchrony. *European Journal of Social Psychology*, *40*, 52–60.

Miles, L. K., Lumsden, J., Richardson, M. J., & Macrae, C. N (2011). Do birds of a feather move together?: Group membership and behavioral synchrony. *Experimental Brain Research*, *211*(3–4), 495–403.

Miles, L. K., Nind, L. K., & Macrae, C. N. (2009). The rhythm of rapport: Interpersonal synchrony and social perception. *Journal of Experimental Social Psychology*, *45*(3), 585–589.

North, A. C., Tarrant, M., & Hargreaves, D. J. (2004). The effects of music on helping behaviour: A field study. *Environment and Behavior*, *36*(2), 266–275.

Petty, R. E., & Cacioppo, J. T. (1979). Issue involvement can increase or decrease persuasion by enhancing message-relevant cognitive responses. *Journal of Personality and Social Psychology*, *37*, 1915–1926.

Rabinowitch, T.-C., & Knafo-Noam, A. (2015). Synchronous rhythmic interaction enhances children's perceived similarity and closeness towards each other. *PLoS ONE*, *10*(4), 1–10.

Rotenberg, K. J., Fox, C., Green, S., Ruderman, L., Slater, K., Stevens, K., & Carlo, G. (2005). Construction and validation of a children's interpersonal trust belief scale. *British Journal of Development Psychology*, *23*, 271–292.

Sanbonmatsu, D. M., & Kardes, F. R. (1988). The effects of physiological arousal on information processing and persuasion. *Journal of Consumer Research*, *15*(3), 379–385.

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences*, *10*, 70–76.

Schwarz, N. (2004). Meta-cognitive experiences in consumer judgment and decision making. *Journal of Consumer Psychology*, *14*, 332–348.

Summerfield, Q. (1992). Lipreading and audio-visual speech perception. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *335*(1273), 71–78.

Tomasello, M., & Carpenter, M. (2007). Shared intentionality. *Developmental Science*, *10*(1), 121–125. Trehub, S. (2003). Musical predispositions in infancy: An update. In R. Zatorre & I. E. Peretz (Eds.), *The cognitive neuroscience of music* (pp. 3–20). Oxford, UK: Oxford University Press.

Trevarthen, C. (1999–2000). Musicality and the intrinsic motive pulse: Evidence from

human psychobiology and infant communication. *Musicae Scientiae*, *3*(1 Suppl.), 155–211.

Valdesolo, P., & DeSteno, D. (2011). Synchrony and the social tuning of compassion. *Emotion*, *11*(2), 262–266.

Valdesolo, P., Ouyang, J., & DeSteno, D. (2010). The rhythm of joint action: Synchrony promotes cooperative ability. *Journal of Experimental Social Psychology*, *46*(4), 693–695.

Wentzel, K. R. (1991). Relations between social competence and academic achievement in early adolescence. *Child Development*, *62*, 1066–1078.

Whittlesea, B. W. (1993). Illusions of familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(6), 1235–1253.

Wiltermuth, S. S., & Heath, C. (2009). Synchrony and cooperation. *Psychological Science*, *20*(1), 1–5. Winkielman, P., & Cacioppo, J. T. (2001). Mind at ease puts a smile on the face: Psychophysiological evidence that processing facilitation elicits positive affect. *Journal of Personality and Social Psychology*,

81, 989–1000.Winkielman, P., Schwarz, N., & Nowak, A. (2002). Affect and processing dynamics. In S. C. Moore & M. Oaksford (Eds.), *Emotional cognition: From brain to behaviour* (pp. 111–135). Amsterdam, the Netherlands: John Benjamins.Winkler, I., Háden, G. P., Ladinig, O., Sziller, I., & Honing, H. (2009). Newborn infants detect the beat in music. *Proceedings of the National Academy of Sciences of the United States of America, 106*(7), 2468-2471.

Woolhouse, M. H., Tidhar, D., & Cross, I. (2016). Effects on interpersonal memory of dancing in time with others. *Frontiers in Psychology, 7*. doi:10.3389/fpsyg.2016.00167

Žonja, S., Livun, N., & Jambrošić, K. (2006, June). Audio-visual interaction: Multimedia applications. In *Forty-eighth International Symposium ELMAR-2006 focused on Multimedia Signal Processing and Communications* (pp. 143–146). New York, NY: IEEE. doi:10.1109/ELMAR.2006.329535Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, *9*, 1–27.