Beep Beep: Building Trust with Sound

Matthias Schmidmaier
LMU Munich
TAWNY
Munich, Germany
matthias.schmidmaier@ifi.lmu.de

Dominik Maurice Runge
LMU Munich
Munich, Germany
dom.runge@campus.lmu.de

Heinrich Hußmann
LMU Munich
Munich, Germany
hussmann@ifi.lmu.de

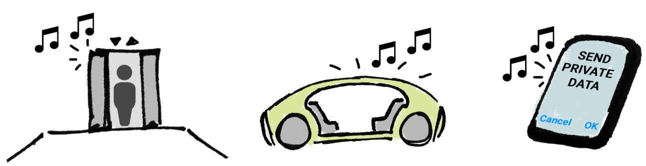

Figure 1: Using sound feedback to create trust in HCI.

CHI '20 Extended Abstracts, April 25–30, 2020, Honolulu, HI, USA.

ACM ISBN 978-1-4503-6819-3/20/04.
http://dx.doi.org/10.1145/3334480.3382848

Abstract
Audio is one modality that besides content transmission offers non-verbal cues that influence emotional perception. This allows to increase trust for example in privacy-sensitive systems like digital assistants. In this work we focus on basic audio feedback and explore how parameters like melody, pitch or tempo influence the creation of trust. We refer to related research in trust perception of voice, and evaluate if the derived concepts can be universally applied to simple sound patterns. Our study (n=39) shows significant effects for melody and mode, while tendencies were found for pitch and individual user preferences. We consider our findings to serve as basis for research towards the design of unobtrusive and trustworthy user experiences.

Author Keywords
Trustful HCI; Audio feedback; Affective Interaction Design.

CCS Concepts
•Human-centered computing → HCI theory, concepts and models; Sound-based input / output;

Introduction
The tone of a voice, but also music and noises can influence emotions, decisions and our assessment of other people, systems and situations. As with interpersonal interaction, interaction with devices and applications requires a certain

Trust Dimensions

Cognitive: trust based on knowledge and experience (longitudinal nature of trust) [6, 14].

Emotional: trust based on emotional relationship between trustor and trustee [14].

Behavioral: trust based on trust-implying actions [14].

Melodie Patterns

Ascending: three ascending notes separated by one octave. Note sequence: *c'-g'-c"*.

Valley-shaped: four notes, first descending then ascending. Note sequence: *c"-e'-g'-c"*.

Arch-shaped: four notes, inverted valley-shaped melody. Note sequence: *c'-g'-e'-c'*.

Descending: inverted version of the ascending melody. Note sequence: *c"-g'-c'*.

amount of trust, depending on context and delegated tasks and responsiblities. People usually trust established systems like elevators and cars but also popular software and services (see Figure 1). In 'first encounter' situations with unknown applications, a user may rely on third-party guarantees. For that, it sometimes is enough to provide simple symbols without giving detailed information: a green lock besides an URL tells us that we (hopefully) enter a trustworthy website. However, besides rational factors like experience, also emotions and system behavior have significant impact on a user's decision to trust and interact [14]. Human-robot interaction and digital assistants make use of anthropomorphic design to express emotional states and simulate human characteristics. In addition, affective computing allows to recognize, adapt to and as a result influence a user's affective state [18]. In this work we explore if principles found for emotion and trust perception through music and voice, apply for simple sound patterns. Our intention is to provide basic research for trustworthy interaction and UX design using the audio modality. We expect our findings to serve as base for context-specific research of non-verbal audio trust cues. We further suggest that audio feedback that is designed regarding trust perception can improve interaction for visually impaired users and in scenarios with exclusively audible feedback.

Related Work

In this chapter we briefly introduce the role of audio and trust in HCI. We further introduce research on emotion expression in music and trustworthiness of voices.

Audio in HCI

Audio is a common modality in HCI, used to transmit verbal information but also to direct attention or emphasize and categorize the source of interaction [2]. In [4], authors highlight the importance of designing sonic interaction not only for functional, signaling purposes but also for 'novel sensory and social experiences' [4]. Similarly, Bramwell-Dicks et al. [2] suggest to use audio for affective interaction.

Trust in HCI

Dictionaries define trust as 'the belief that someone is good and honest and will not harm you, or that something is safe and reliable' [19] or the 'firm belief in the reliability, truth, or ability of someone or something' [20]. Psychologists perceive trust as 'an attitude or intention' [21], while in sociology trust is defined as 'a reflection of behaviors, choices and decisions' [21]. In computer science in turn, trust may be seen as a user's confidence into and willingness to interact with a system [21]. Lee and See [13] view trust in the context of automation as the attitude that an agent (trustee) will help an individual (trustor) achieving their goals in uncertain or vulnerable situations. Finally, according to Lewis et al. [14] in a sociological context, trust can be divided into a cognitive, emotional and behavioral dimension (see sidebar 'Trust Dimensions'). Concepts of interpersonal trust may be also applied to HCI [11]. Based on related research [15, 9], Häuslschmid et al. [10] divide trust-influencing features of technological systems into two groups 'system performance' and 'emotional aspects'. Hoff et al. [11] describe different factors that influence trust in automated systems. They relate to various research [7, 8, 17] that suggests anthropomorphism of interfaces to establish trust. Further factors are high usability, polite system communication, system transparency and enabling user intervention [10].

Emotion through Sound

In [5], authors describe how musical cues (tempo, dynamics, mode, articulation, timbre, phrasing) influence the perception of emotions. Further research showed that tempo, pitch and mode are the most relevant cues for the expression of happiness (fast, high pitch, major mode) and sadness (slow, low

	tune	p	m	t	notes
1.	asc	▽	C	+	
2.	asc	▲	C	+	
3.	asc	▽	**Cm**	+	
4.	asc	▽	C	-	
5.	valley	▽	C	+	
6.	valley	▲	C	+	
7.	valley	▽	**Cm**	+	
8.	valley	▽	C	-	
9.	arch	▽	C	+	
10.	arch	▲	C	+	
11.	arch	▽	**Cm**	+	
12.	arch	▽	C	-	
13.	desc	▽	C	+	
14.	desc	▲	C	+	
15.	desc	▽	**Cm**	+	
16.	desc	▽	C	-	
17.					
18.					
19.	test	▽	C	+	

Table 1: Ascending, valley-, arch-shaped and descending melodies. Varying in pitch (high ▲, low ▽), mode (C/*Cm*) and tempo (fast +, slow -). Patterns 17, 18 and 19 describe Android, iOS and test melodies.

pitch, minor mode) [3]. Moving from music to sound feedback in HCI - more specific HRI - Jee et al. [12] explored how social robots could use simple sounds to express intention (affirmation, denial, encouragement, introduction, question) and emotions (happiness and sadness), identifying intonation, pitch and timbre as dominant parameters. Authors further suggest anthropomorphic sound design (pitch and intonation similar to human voice) and infer that timbre can be used to create the perception of personal characteristics like age, gender or honesty [12].

Trustful Voices

Voice is an important modality for personality assessment in social interaction, especially for 'first impressions' and in non-visual scenarios. McAller et al. [16] investigated vocal perception of brief utterances in the two dimensions valence and dominance, defining valence as corresponding to likeability, trustworthiness and warmth. They investigated correlations between eight audio measures. Results indicate different correlations for male and female voices: perceived positive valence is linked to higher average pitch in male voices and rising intonation in female voices [16]. In a subsequent work, the authors explored trustworthiness of computer-generated voice patterns, confirming the influence of pitch and intonation: perceived trustworthiness was found to be low for patterns with low pitch and flat or slightly rising intonation. It increased with rising pitch and for valley-shaped, first falling then rising signals [1]. We used the findings of this research as basis of our sound design.

Sound Pattern Design

In our study we evaluated the influence of different audio parameters on perceived trust. For that we designed various audio patterns, each of which differing to a defined base pattern in one parameter like for example pitch.

Conception

Based on research in emotion expression through music [5, 3] and in trustworthiness of voice [16, 1], we explored four audio parameters: melody, pitch, mode and tempo. In voice, intonation was found to influence valence [16] as well as the perception of trust [1]. We transferred these findings from voice to sound patterns and designed **four melodies** with notes in (1) ascending, (2) valley-shaped, (3) arch-shaped and (4) descending order (see sidebar 'Melodie Patterns'). **Pitch** is important for emotional perception, both in music and in voice. Higher notes were found to be related to happiness [3] and high trustworthiness [1]. Based on the 440Hz standard tuning, our low frequent patterns use frequencies between 261Hz and 523Hz while the high frequent versions are designed one octave higher between 523Hz and 1046Hz. Musical **modes** can be used to express happiness (major) and sadness (minor) [3]. We produced each melody in major and minor scale to test a possible link between mode and trust perception. As fourth parameter we investigated **tempo**, choosing two variations: 140ms per tone (~428bpm) as low tempo and 170ms (~353bpm) as fast tempo, leading to playback durations of 420ms for the shortest and 680ms for longest patterns. As McAller et al. [16] suggest that 390ms is enough time to tell if a voice ist trustworthy, we assume the same for our experiment. Preliminary tests with synthetically created sounds revealed that it is hard for users to rate the trustworthiness of a sound if they cannot identify themselves with the overall **sound style**. We therefore produced all patterns in double versions, once as synthetic, once as piano sound.

Audio Production

We produced 16 audio patterns that vary in melody, mode, pitch and tempo and three additional samples (see table 1). Sample 17 simulates an *Android* melody, while sample 18 simulates a standard *iOS* message sound. With these sam-

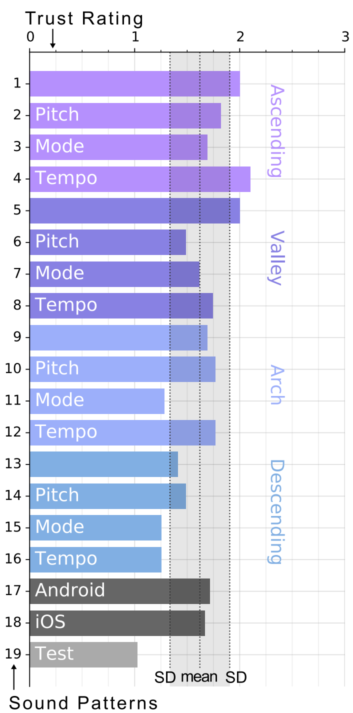

Figure 2: Mean trust ratings for all 19 sound patterns. Grouped by melodies with each group containing a variant with high pitch, C-minor mode and low tempo. Trust rating range: $0..3$.

ples we intended to explore correlations between familiar audio patterns and trustworthiness. Sample 19 acts as test pattern to cross-check the sound style preferences in our study. All samples were mastered with normalized volume level and exported in 44100Hz stereo with a 512-point-sync resampling quality. You can find all patterns on https://chapek9.org/trustsounds. For synthetic sound production we used a wavetable synthesizer and to create a typical 'device feedback' sound. We chose a fast attack of 3ms with no hold but a direct and relative fast decay of 392ms, a sustain value of around -6dB and nearly no release curve (94ms). These settings create short staccato sounds, similar to short piano key strikes, accomplishing close comparability between our two sound styles. Additionally, we applied a small high pass filter in combination with a compressor. The piano sounds were generated with prerecorded audio samples. By duplicating the synth patterns we guaranteed that melody, length, mode and frequency are identical to the synth patterns. We applied a weak compressor and equalizer to achieve a strong and well balanced sound.

Experiment

We conducted an online study with n=39 subjects, inspired by the experiment exectued by Belin et al. [1]. The study URL was distributed via personal channels. From finally 40 participants 39 finished the study, while one user canceled and therefore was ignored for evaluation.

Design and Procedure

We designed our online study as sequential website, starting with an introduction and terms of agreement about privacy and data processing. Subsequently, we asked for demographic data, preferred mobile operating system (OS) and favorite music genre. Before starting the audio part, users were advised to put on headphones if possible and turn the volume up. For sound style choice, users were asked to listen to a piano and a synthetic sample and select which they prefer most. Based on this selection, the corresponding 19 sounds were presented to the user successively in randomized partial counterbalanced order, achieving a within-subjects design. To create an understanding of trustworthy feedback, subjects were asked to consider if they ‚like' a sound and if they are ‚feeling good' while hearing it. Further they were asked to imagine how the sound feedback would influence their interaction with a system in a trust context like entering an elevator. After each playback, the user got to rate the trustworthiness of the pattern on a 4-point rating scale, answering the question 'I found this audio pattern trustworthy' choosing from 'strongly disagree', 'disagree', 'agree' and 'strongly agree'. We did not provide a neutral rating option as in trust scenarios a user usually also has to make a decision and is not able to stay in between. The overall approach is similar to the one used in the voice study by Belin et al. [1]. Finally, participants were asked to take a break halfway to ensure concentration and counter musical insensitivity.

Hypotheses

Regarding the influence of different audio features on perceived trustworthiness we tested five hypotheses. **[H1]** Ascending melodies are perceived as more trustworthy than descending ones: as raising intonation and ascending melodies in music are linked to positive valence, we expected that the same holds true for trust perception in sound patterns. **[H2]** Valley-shaped melodies are perceived as more trustworthy than simply ascending melodies: as voice with valley-shaped intonation is perceived more trustworthy, we assumed the same for first descending, then ascending melodies. **[H3]** Higher pitch results in higher perceived trustworthiness: again, the same effect was found for human voices. **[H4]** Major mode audio patterns are perceived as more trustworthy than patterns in minor mode: we assumed that the influence of musical modes on the perception of trustworthiness corre-

	base	*valley*	*arch*	*desc*
♪	1,2 3,4	5,6 7,8	9,10 11,12	13,14 15,16
$\bar{x}$	1.90	**-0.19***	**-0.27***	**-0.55***
♂	1.94	**-0.28***	**-0.45***	**-0.63***
♀	1.84	-0.11	+0.04	**-0.53***
	1.99	**-0.25***	**-0.49***	**-0.74***
🤖	1.82	-0.14	-0.11	**-0.45***
p	1.88	-0.09	**-0.30***	**-0.63***
s	1.93	**-0.30***	-0.25	**-0.47***

Table 2: Rating deviations for melodies.[1]

	base	*pitch*	*mode*	*tempo*
♪	1,5 9,13	2,6 10,14	3,7 11,15	4,8 12,16
$\bar{x}$	1.78	-0.14	**-0.32***	-0.06
♂	1.68	+0.01	**-0.25***	-0.09
♀	1.91	**-0.35***	**-0.41***	-0.02
	1.74	-0.13	**-0.38***	+0.2
🤖	1.79	-0.12	**-0.28***	**-0.18***
p	1.71	-0.01	**-0.30***	-0.05
s	1.84	**-0.26***	**-0.33***	-0.06

Table 3: Rating deviations for pitch, mode and tempo.[1]

[1] Tables show rating deviations from baseline (*base* column) for male ♂ and female ♀ participants, iOS and Android 🤖 users and groups that preferred piano (*p*) or synthetic (*s*) sound style. Ratings range from $0..3$, with significant deviations **highlighted***. First row shows grouped sound patterns ♪.

lates with the influence on emotion expression. **[H5]** Faster tempo results in a higher level of perceived trustworthiness: as faster melodies in music are perceived as more attractive and friendly, we assumed the same regarding trustworthiness.

Results

Participants were aged between 18 and 64 years, with 22 male, 16 female and one preferring not to specify gender. Regarding preferred mobile operating system, 19 users selected 'Android', 18 participants 'iOS' (two 'other'). Sound style selection was also balanced, with 20 participants selecting piano and 19 selecting synthetic style. Selection of preferred genre is distributed between Alternative (9), Pop (8), Rock (5), Electro (5), Metal (3), Hip-Hop (3), Chillout (0), Country (0) and 'I am individual' (5). For evaluation we use a numeric representation of the rating scale ranging from 0 for 'strongly disagree' to 3 for 'strongly agree'. Figure 2 shows the overall mean trust rating for each sound pattern. Participants rated the audio patterns with an overall $mean = 1.62$ ($SD = 0.28$). Patterns 1 to 16 were rated with an overall $mean = 1.65$ ($SD = 0.27$).

Melodies A Friedman test ($\alpha = 0.05$) and pairwise Wilcoxon post-hoc analysis ($\alpha = 0.05$) showed significant differences between all melodies, except for the comparison between valley-shaped and arch-shaped melodies. Table 2 shows means and differences of trust rating for melodies compared to the baseline melody (ascending; in patterns 1 to 4). While male participants aligned with the overall ratings, for female participants significant differences to the baseline could only be found for descending melodies. Similarly, for iOS users significant differences aligned with overall ratings while for Android users significant differences could only be found for descending melodies. Results showed significant differences for the arch-shaped as well as descending melodies for participants that chose piano style. The choice of synthetic sound style resulted in significant differences for valley-shaped and descending melodies.

Pitch, Mode and Tempo We further ran pairwise Wilcoxon tests (α=0.05) between pitch, mode and tempo and the corresponding baseline patterns. While the test revealed a significant difference for mode, no significance was found for pitch and tempo. While the same accounts for male participants, for female participants we found also significant differences in pitch. Table 3 shows means and differences for these parameters. For Android users, we found significant differences for parameters mode and tempo. Furthermore, for participants that chose synthetic sound style, results showed significant differences for pitch.

I am Individual Regarding preferred music genre, group sizes were found to be too small for reliable statistical evaluation. Nevertheless, the results provided interesting tendencies. Participants who selected Hip Hop as prefered genre rated minor mode patterns higher ($mean = 2.0$) compared to the corresponding baseline ($mean = 1.67$), contradicting the general rating behavior. The 'Android pattern' was rated with an overall $mean = 1.72$, almost the same as the 'iOS pattern' with $mean = 1.67$. Android users rated the Android pattern with $mean = 1.74$, the iOS pattern with $mean = 1.79$. iOS users in turn rated the Android pattern with $mean = 1.67$, the iOS pattern with $mean = 1.50$. However, participants that selected the 'I am individual' option on genre rated the Android pattern low ($mean = 1.40$), but the iOS pattern as most trustworthy pattern overall ($mean = 2.0$). Finally, the test pattern - a sound in piano style for participants who chose synthetic and vice versa - was rated with a comparably low $mean = 1.03$.

	valley	arch	desc
coef	**-0.19***	**-0.28***	**-0.55***
	pitch	*mode*	*tempo*
coef	+0.13	**-0.31***	-0.06
	gender	*sound*	*OS*
coef	+0.10	-0.03	-0.05

Table 4: OLS coefficients for the influence of audio and user features on trust rating. Significant features are **highlighted***. Trust rating range: $0..3$.

Linear Regression To get additional insight into the correlative influence of audio and user features, we performed a linear regression (Ordinary Least Squares, with intercept) over ratings for patterns 1 to 16, using pattern 1 as baseline. Regression results ($\alpha = 0.05$) as seen in Table 4 showed significant influence of the parameters 'valley-shaped', 'arch-shaped', 'descending' and 'minor mode'.

Discussion

Our results showed significant influence of melody patterns on perceived trustworthiness. The greatest difference was found between purely ascending and descending melodies. Also the valley- and arch-shaped melodies showed significant differences to ascending and descending melodies, yet not between each other. Based on these findings, we accepted *H1*, however had to reject *H2*. Regarding pitch, we had to reject our hypothesis *H3* for an overall view. The separate views on ratings of female participants and participants who selected the synthetic sound style were the only ones showing significance for pitch. However, linear regression did not show significant influence of gender and sound style selection. We still suggest to not completely ignore these factors in future work due to findings in related research. The pitch difference in our patterns only was one octave. We suggest to explore wider ranges in frequency in future research. Hypothesis *H4* could be clearly supported: the level of perceived trustworthiness was significantly lower for minor mode patterns compared to major mode. Yet, again the limited variable range we used should be considered as we only compared C-major and C-minor patterns. Furthermore, tendencies found in the context of preferred music genre let us assume that there might be potential for personalization. The influence of tempo on trust perception was not found to be significant. We therefore rejected hypothesis *H5*. Rating for iOS and Android sound patterns were not found to be significantly related to iOS and Android users.

Limitations Although our results showed significant differences that partly confirmed findings of related work, we are aware of certain limitations. For one thing, the number of participants ($n = 39$) was rather low compared to related studies [1]. Second, in pattern design we found it difficult to choose appropriate thresholds, for example: is one octave difference in pitch enough? Especially in tempo, we suggest to increase the gap between values for 'slow' and 'fast' patterns in future research. In addition, our patterns were not designed regarding the transmission of explicit information like 'alert' or 'confirmation'. Design rules for this 'verbal' part of feedback would also set limitations compared to purely trustworthy design.

Conclusion

Audio is one modality that allows swift and non-intrusive interaction in combination with multimodal feedback but also as primary feedback for visualy impaired users. In this paper we explored if and which audio parameters in short sound patterns influence the perception of trust. From our results we conclude that specific audio parameters like melody or mode can be used for trustworthy interaction design. Our findings also indicate possible correlations between individual user characteristics and audio parameters aside from mode and melody. For future work, we suggest to take personalization and adaptive interface design into account. In this context we also suggest to research correlations with a user's cultural background and - together with individual music and preferences - take that into account for feedback design. Besides to 'static' user features like preferences, applications could react dynamically to a user's affective state and context: 'uncertainty detected, increasing trust cues in communication'. In this context, machine learning approaches could be used to explore and adapt individual factors and thresholds for trust perception.

REFERENCES

[1] Pascal Belin, Bibi Boehme, and Phil McAleer. 2017. The sound of trustworthiness: Acoustic-based modulation of perceived voice personality. *PLoS One* 12, 10 (Oct. 2017), e0185651.

[2] Anna Bramwell-Dicks, Helen Petrie, Alistair D N Edwards, and Christopher Power. 2013. Affective Musical Interaction: Influencing Users' Behaviour and Experiences with Music. In *Music and Human-Computer Interaction*, Simon Holland, Katie Wilkie, Paul Mulholland, and Allan Seago (Eds.). Springer London, London, 67–83.

[3] Tuomas Eerola, Anders Friberg, and Roberto Bresin. 2013. Emotional expression in music: contribution, linearity, and additivity of primary musical cues. *Front. Psychol.* 4 (July 2013), 487.

[4] Karmen Franinović and Stefania Serafin. 2013. *Sonic Interaction Design*. MIT Press.

[5] Alf Gabrielsson and Erik Lindström. 2010. The role of structure in the musical expression of emotions. *Handbook of music and emotion: Theory, research, applications.* 975 (2010), 367–400.

[6] David Gefen, Izak Benbasat, and Paul A Pavlou. 2008. A Research Agenda for Trust in Online Environments. *Journal of Management Information Systems* 24, 4 (2008), 275–286.

[7] Li Gong. 2008. How social is social responses to computers? The function of the degree of anthropomorphism in computer representations. *Comput. Human Behav.* 24, 4 (July 2008), 1494–1509.

[8] Brian Daniel Green. 2010. *Applying human characteristics of trust to animated anthropomorphic software agents*. State University of NY at Buffalo.

[9] D Harrison McKnight and Norman L Chervany. 2001. Trust and Distrust Definitions: One Bite at a Time. In *Trust in Cyber-societies*. Springer Berlin Heidelberg.

[10] Renate Häuslschmid, Max von Bülow, Bastian Pfleging, and Andreas Butz. 2017. SupportingTrust in Autonomous Driving. In *Proceedings of the 22Nd International Conference on Intelligent User Interfaces (IUI '17)*. ACM, New York, NY, USA, 319–329.

[11] Kevin Anthony Hoff and Masooda Bashir. 2015. Trust in automation: integrating empirical evidence on factors that influence trust. *Hum. Factors* 57, 3 (May 2015), 407–434.

[12] Eun-Sook Jee, Yong-Jeon Jeong, Chong Hui Kim, and Hisato Kobayashi. 2010. Sound design for emotion and intention expression of socially interactive robots. *Intelligent Service Robotics* 3, 3 (July 2010), 199–206.

[13] John D Lee and Katrina A See. 2004. Trust in automation: designing for appropriate reliance. *Hum. Factors* 46, 1 (2004), 50–80.

[14] J David Lewis and Andrew Weigert. 1985. Trust as a Social Reality. *Soc. Forces* 63, 4 (1985), 967–985.

[15] Roger C Mayer, James H Davis, and F David Schoorman. 1995. An Integrative Model Of Organizational Trust. *AMRO* 20, 3 (July 1995).

[16] Phil McAleer, Alexander Todorov, and Pascal Belin. 2014. How do you say 'hello'? Personality impressions from brief novel voices. *PLoS One* 9, 3 (March 2014).

[17] Richard Pak, Nicole Fink, Margaux Price, Brock Bass, and Lindsay Sturre. 2012. Decision support aids with anthropomorphic characteristics influence trust and performance in younger and older adults. *Ergonomics* 55, 9 (July 2012), 1059–1072.

[18] Rosalind Wright Picard and Others. 1995. Affective computing. (1995).

[19] Cambridge University Press. 2019a. *Cambridge Dictionary*. https://dictionary.cambridge.org/

[20] Oxford University Press. 2019b. *Oxford English Dictionary*. https://dictionary.cambridge.org/

[21] Sonia Sousa, David Lamas, and Paulo Dias. 2014. A Model for Human-Computer Trust. In *Learning and Collaboration Technologies. Designing and Developing Novel Learning Experiences*. 128–137.