The Ethics of Online Video Analysis for Systematic Observation of Behavior

Malte F. Jung
Center for Design Research
Stanford University
mjung@stanford.edu

Steven P. Dow
HCI Group
Stanford University
spdow@stanford.edu

ABSTRACT
Social scientists have gained powerful insights, often countering established theory, through systematic observation of behavior in video. However, video analysis techniques are resource intensive, requiring human investment to design, and execute a coding scheme. Moreover, the independence of coders is hard to guarantee when multiple coders work in the same laboratory where they talk about their work and thus inadvertently influence each other's observations. Crowd-sourcing platforms provide an opportunity to distribute the coding task to distributed online workers. With this opportunity comes a set of new ethical challenges mainly around the difficulties in protecting subject's privacy. In this position paper we start to unpack these challenges such as inadvertent release of video, misjudgments of the risks involved, and the intrusiveness of different coding schemes. We introduce thin-slicing, coder-tracking, and video distortion as possibilities for risk mitigation.

Author Keywords
Systematic observation of behavior, video coding, ethical challenges, thin slicing.

ACM Classification Keywords
J.4 Social and Behavioral Sciences; K.4.1 Public Policy Issues: Ethics; H.5 Information Interfaces and Presentation (e.g. HCI)

INTRODUCTION
A goal of many social scientists is to understand what people actually do and how their activities unfold over time [4, 11, 12, 14, 18]. Video recording and analysis is an integral technique for social scientists to systematically observe and code human behavior over time. Video methods yield insights about participant dynamics that cannot be understood through self-report measures alone [18]. Video-based systematic observation of behavior has been adopted to study negotiations [18], marital interactions [12], cancer support group interactions [11], and pair programming interactions [14], to name a few. One illustrative example is Gottman and Levinson's observation of married couples [13]. They were able to accurately predict (up to 93%) whether a couple will divorce years later by observing short video clips of couples in conflict. They coded 15-minute "thin slices" of interaction for a specific set of emotion-related behaviors. Even a three-minute glimpse was shown to yield predictive power [5]. In our research we use this method to systematically observe affective behavior in a thin slice of interactions of design teams and relate these observations to performance-relevant outcomes [14].

CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.

AN ONLINE VIDEO ANALYSIS SERVICE
We are currently designing and developing a crowd-based video analysis service. Crowdsourcing platforms like Amazon Mechanical Turk [1] provide an opportunity to leverage large pools of human observers. We hypothesize that video coding tasks can be distributed and effectively analyzed by a larger community of researchers at reduced cost. Coding time can also be reduced through automated coding, and state of the art computer vision now allows automated coding of what Bakeman and Gottman call physically based coding schemes such as the Facial Action Coding System [4, 7, 9, 16]. However, such automated coding techniques cannot classify behavior according to highly contextual socially based schemes such as SPAFF [6]. Human observers, however, can flexibly adapt to different coding schemes. Even untrained observers can generate reliable results on complex socially based coding tasks [17].

With our service we want to address three major challenges that are currently acting as a barrier towards a broader adoption of this method.

First, the current methods to generate data using systematic observation of behavior are highly time consuming. Coding only a few minutes of video using the marital interaction coding scheme can take several hours [11]. A more tedious coding scheme like the Facial Action Coding System [9] can take even longer. In addition to the coding time, these observation schemes require extensive training to produce reliable results.

Second, obtaining high observer reliability requires elaborate procedures such as careful training, continuous calibration, and frequent test of coder reliability [4]. When doing coding in a collocated setting it can be difficult to prevent observers from influencing each other. Ideally, coders have to be kept blind to hypotheses but often this goal is hard to maintain especially when coders are working over longer durations with a lab and are integrated into other research tasks.

Third, it takes time and effort to build and maintain an infrastructure for video coding. A coding setup needs dedicated space, coders have to be recruited, continuously supervised, and scheduled. Additionally a social environment should be provided that supports coders to maintain motivated with the often tedious and boring task of coding video. These infrastructure challenges led researchers to invest into dedicated "coding labs" like the one at the University of Calgary [10].

Our envisioned service will segment and securely display video online to crowd workers, elicit judgments or tallies using researcher-defined features, and reliably compile the results. The service will have three main components: a Web interface for researchers to upload video and specify coding parameters, a backend routine for segmenting video and compiling results, and worker task pages for collecting judgments from crowdsourcing platforms such as Amazon Mechanical Turk. Worker redundancy, filtering and calibration tasks will ensure more accurate coding. The service will not only provide practical value, researchers can ground theoretical claims on impartial independent analysis.

In parallel with the development of a video coding service we are designing a study in which we want to use the aforementioned service to obtain behavioral data from videotaped interactions. The aim of our study is to re-investigate a set of videos that were recorded during a study on prototyping interactions [8] and relate interaction characteristics to performance relevant outcomes. This study in conjunction with the workshop provides an opportunity to discuss the important ethical questions early in the development of our service.

RISKS OF ONLINE VIDEO CODING

The possibility of distributed video analysis introduces a new set of ethical challenges. A release of video for online coding runs at risk of violating many of the guidelines proposed earlier [15]. How can we manage the challenge of protecting participant identity while maintaining a high quality of data? There are not only risks for the subjects featured in the videos but also for those who are asked to code them. The development of a successful video analysis service requires a careful analysis of the risks involved. We identified an initial set of risks that we are currently addressing in our development.

Coders may inadvertently release a video clip

One of the major risks in an online distribution of video for coding is an inadvertent release of a video. As soon as a video is published online, its use can no longer be controlled. Even with sophisticated copy protection mechanisms, it is still possible to just record a video off a screen with a camera. A release of video can lead to a series of scenarios that can have severe implications for the psychological well-being of a study participant [15]. One of the major questions here is whether the content of video affects the decision about whether it can be distributed or not. For example, video coding has been used previously in studying emotion expression in breast cancer support groups [11]. Clearly, such content is much more sensitive, and the subjects are much more vulnerable than those recorded during, for example, a public presentation.

Participants may be unaware of risks involved

In preparing our study we were surprised how many of the participants we asked were readily agreeing to their video being used for an online evaluation. It is possible that many participants were not aware of the actual risks this permission involved. Additional current human subjects approval boards seem not well prepared to evaluate studies that use this kind of crowd based video analysis. This poses the question how we can communicate the risks in a way that allow an accurate assessment of the risks involved.

Some judgments are more intrusive than others

The type of coding might affect how vulnerable subjects feel about their video being rated. An evaluation of a person's emotion might be perceived as far more intrusive than an evaluation of how often they picked up a pen for example. The question therefore arises whether subjects have to be told, according to what criteria their video will be evaluated, and whether a permission to use a video should be limited to the use of a particular coding scheme.

Coders may identify participants

Finally, there is the challenge of protecting the identity of human subjects while providing high quality video. Since it is difficult to control who signs up as a coder, there is the risk that subjects get identified by coders and sensitive information might be shared.

RISK-MITIGATION AND TRADEOFFS

Given the risks listed above, we want to propose a set of initial techniques that might help in alleviating them. Each of these techniques comes with its own set of tradeoffs and each possibility can only be a starting point for a discussion about risks and possibilities in dealing with them.

Thin-slicing

One possibility to protect a participant's identity could be to take only a short sample or "thin slice" for coding. A coder would never gain access to more than a few seconds of video. The thin-slicing technique refers to the process of making accurate classifications based on small samples, or "thin slices" of expressive behaviors [2, 3]. The thin-slicing research showed powerfully that certain behavioral characteristics are stable over time and that only a small interaction sample is necessary to make meaningful judgments about behavior occurring over longer durations such as hours, or even months. In a meta-analysis across 38 different studies, Ambady and Rosenthal [2] were able to show that short behavioral samples ranging between 20 seconds and 5 minutes, are highly indicative of long-term characteristics, irrespective of the specific context they were taken in.

Potential downsides for using thin slicing are a loss of context. This can not only influence the quality of coding but also some statements by subjects, when taken out of context, can be more damaging when released. Additionally thin slicing per se does not protect a subject from being identifiable.

Tracking coders

Another possibility to improve the protection of subject's privacy might be to recruit only coders who are willing to self-identify. Coder identification would allow to track access of videos by their respective coders. Embedding invisible "watermarks" into the videos could make it possible to trace an inadvertently released video back to the coder who violated a subject's privacy.

The downside of this approach is that damage cannot be prevented. Knowing that videos are traceable might inhibit coders from sharing videos, but ultimately a release cannot be avoided.

Distorting audio and video data

Finally it is possible to distort the audio and video-tracks partially so that it makes an identification of subjects impossible. This capability could even be built into the envisioned service itself.

The disadvantages of this approach are that it could compromise the quality of the coding. Many coding schemes require high quality video recordings that make each individual readily identifiable. For example the Facial Action Coding System [9], requires a detailed visibility of the face when coding minutes changes in facial muscle configuration. Gottman's Specific Affect Coding System [6] relies on high quality audio for accurate assessments.

CONCLUSION

We introduced crowdsourcing as a technique to leverage systematic observation of behavior for an extended community of researchers and proposed an online video coding service to implement it. We see potential that our video coding service could not only be a resource for coding large amounts of data reliably and fast but also provide a platform for rapidly prototyping and testing new coding schemes. With this technique come new ethical challenges that have to be mastered for it to become a viable alternative to other methods. We discussed specific risks and possible techniques for their mitigation.

We are currently building the proposed mitigation features into our online video coding service. None of these techniques have been tested thoroughly yet and thus researchers aiming to do online video coding should be particularly careful and exercise discretion when placing video online. It is especially important that participants understand exactly how their video will be analyzed. Additionally, the techniques proposed here are technical in nature, which begs the question about a possibility of socio-technical solutions for risk mitigation. For example, would it be possible to recruit a vetted pool of coders? How can ethical guidelines be reinforced socially?

Given the possibility of a distributed human video coding service, it is especially timely to discuss the issues raised here with other researchers facing similar challenges. Crowdsourcing techniques are becoming an increasingly important aspect of general research practice. A set of concrete guidelines will provide direction for designers and researchers.

REFERENCES

1. *Amazon Mechanical Turk*. Available from: www.mturk.com.
2. Ambady, N. and R. Rosenthal, *Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis.* Psychological bulletin, 1992. **111**(2): p. 256-274.
3. Ambady, N. and R. Rosenthal, *Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness.* Journal of personality and social psychology, 1993. **64**(3): p. 431-441.
4. Bakeman, R. and J. Gottman, *Observing interaction: An introduction to sequential analysis*. 1986, Cambridge, UK: Cambridge Univ Press.
5. Carrere, S. and J.M. Gottman, *Predicting divorce among newlyweds from the first three minutes of a marital conflict discussion.* Fam Process, 1999. **38**(3): p. 293-301.
6. Coan, J.A. and J.M. Cottman, *The Specific Affect Coding System (SPAFF)*, in *Handbook of Emotion Elicitation and Assessment*, J.A. Coan and J.J.B. Allen, Editors. 2007, Oxford University Press: New York, USA. p. 267-285.
7. Cohn, J., et al., *Automated face analysis by feature point tracking has high concurrent validity with manual FACS coding.* Psychophysiology, 1999. **36**(1): p. 35-43.
8. Dow, S.P., et al. *Prototyping Dynamics: Sharing Multiple Designs Improves Exploration, Group Rapport, and Results*. in *Conference on Human Factors in Computing Systems (CHI'11)*. 2011. Vancouver, BC, CAN.
9. Ekman, P. and W.V. Friesen, *Facial Action Coding System: A Technique for the Measurement*

of Facial Movement. 1978, Consulting Psychologists Press, Palo Alto, CA.

10. Giese-Davis, J. *The Coding Lab at University of Calgary*. 2005; Available from: http://coding-lab.giese-davis.com/index.php.
11. Giese-Davis, J., et al., *Emotional expression and diurnal cortisol slope in women with metastatic breast cancer in supportive-expressive group therapy: A preliminary study.* Biological Psychology, 2006. **73**(2): p. 190-198.
12. Gottman, J.M. and R.W. Levenson, *Marital processes predictive of later dissolution: Behavior, physiology, and health.* Journal of Personality and Social Psychology, 1992. **63**(2): p. 221-233.
13. Gottman, J.M. and R.W. Levenson, *The Timing of Divorce: Predicting When a Couple Will Divorce Over a 14-Year Period.* Journal of Marriage and the Family, 2000. **62**(3): p. 737-745.
14. Jung, M., J. Chong, and L. Leifer, *Pair Programming Performance: An emotional dynamics point of view from marital pair counseling*. 2010, Electronic Colloquium on Design Thinking Research (ECDTR): Stanford, CA.
15. Mackay, W.E. *Ethics, Lies and Videotape...* in *CHI'95*. 1995. Denver, USA: ACM Press/Addison-Wesley Publishing Co.
16. Tian, Y., T. Kanade, and J. Cohn, *Facial expression analysis.* Handbook of face recognition, 2005: p. 247-275.
17. Waldinger, R.J., et al., *Reading Others' Emotions: The Role of Intuitive Judgments in Predicting Marital Satisfaction, Quality, and Stability.* Journal of Family Psychology, 2004. **18**(1): p. 58-71.
18. Weingart, L.R., M. Olekalns, and P.L. Smith, *Quantitative Coding of Negotiation Behavior.* International Negotiation, 2004. **9**: p. 441-456.