Comparing Social Science and Computer Science

Workflow Processes for Studying Group Interactions

Abstract

In this article, a team of authors from the Geeks and Groupies workshop, in Leiden, Netherlands, compare prototypical approaches to studying group interaction in social science and computer science disciplines, which we call workflows. To help social and computer science scholars understand and manage these differences, we organize workflow into three major stages: research design, data collection, and analysis. For each stage, we offer a brief overview on how scholars from each discipline work. We then compare those approaches and identify potential synergies and challenges. We conclude our paper by discussing potential directions for more integrated and mutually beneficial collaboration that go beyond the producer-consumer model.

Keywords: research methods, data analysis, interdisciplinary collaboration

At its core, research on group interaction is interdisciplinary (Wagner et al., 2011) and draws on many different research traditions and practices. At a workshop in Leiden, Netherlands, a group of social scientists and computer scientists gathered to discuss potential for collaboration across disciplines to move the study of groups and teams forward for both disciplines in a multi-faceted and interdisciplinary manner. It was quickly realized through these initial interactions that in order to deepen collaboration between social scientists and computer scientists interested in group interaction, researchers must go beyond shared research interests and understand the nuts and bolts of each other's research processes. In fact, constructive conflict quickly arose during the workshop concerning how each discipline works. We call these issues *workflows*, which encompass the research design, data collection, data analysis, and dissemination of results. In particular, we compare prototypical workflow approaches to studying group interaction in each discipline. We aim to provide a starter guide for researchers in both disciplines interested in collaboration. In doing so, we make many generalizations about each discipline and necessarily minimize differences within each discipline, exaggerating their homogeneity. We provide references to more comprehensive reviews of research methods so that the readers may read more about the topics mentioned.

Social Science and Computer Science Workflow

Imagine that thousands of videos of group interaction were available for both social scientists and computer scientists to use simultaneously. In the social sciences, groups and teams scholars have collected video and audio data from thousands of group and team interactions and annotated them (i.e., coded/labeled them) for specific behaviors and states (e.g., McGrath & Altermatt, 2001; Waller & Kaplan, 2016). Meanwhile, computer scientists studying group interaction use a variety of techniques to measure and induce observable correlates of many

similar behaviors and states in group interactions (e.g., Gatica-Perez, 2009), though they also tend to develop technologies to interpret individual or dyadic behaviors (e.g., Vinciarelli et al., 2012). The potential for synergy should be obvious. However, if only computer scientists could access and analyze the many thousands of annotated videos that social scientists, or groups researchers, have produced to provide a broader base for developing more sophisticated automated analytic techniques. If only groups researchers could access and use the computer scientists' approaches to automate coding of the behaviors or emotions of group members, rather than going through the laborious process of coding them by hand.

This producer-user model is an obvious point of potential synergy, and yet may be only the beginning of deeper collaborations between social and computer scientists studying group and team interactions. In order for deeper collaborations to emerge, the many barriers to collaboration between group interaction scholars from different disciplinary backgrounds need addressing. Even if scholars agree on a shared research topic, they must still contend with differences in workflows--the processes by which research is conducted. Scholars from different disciplines may have different approaches to and terminology for designing, collecting, and analyzing data on group interactions, and they may work on very different time scales for collecting, analyzing, and disseminating research results (Wagner et al., 2011). Such differences can inhibit communication between disciplines and prevent potential collaborations from being explored.

To help social and computer science scholars understand and manage these differences, we organize workflow into three major stages: research design, data collection, and analysis. For each stage, we first offer a brief overview on how scholars from each discipline work. We then compare these approaches and identify potential synergies and challenges. Finally, we discuss

potential directions for more integrated and mutually beneficial collaborations beyond the producer-consumer model.

Social Scientist Workflow

At a basic level, most social science scholars aim to publish their research in peer-reviewed journals. The publication process is a long one – top journals routinely take two to four months to return manuscripts after submission, request two or three revisions (each requiring several months of authors' work), and may or may not accept them for publication, so rejected manuscripts must be resubmitted to another journal. Thus, the time from initial submission to final acceptance of a peer-reviewed article takes around a year, often longer. Although social scientists present their research at peer-reviewed conferences and these conferences sometimes publish proceedings, these have lower status and are rarely cited by other researchers. The time-consuming work of publishing is in addition to the process of research design, data collection, and analysis below.

Research design: Deductive or inductive. For groups and teams scholars from the social sciences, researchers usually propose one or more specific relationships to study (hypotheses)*,* between two or more theoretical concepts (variables)*.* Hypotheses typically specify causal relationships between the variables, although they can also predict positive or negative associations or investigate conditions under which such associations are stronger or weaker (Mohr, 1982). This approach is *deductive--*using existing theories and empirical findings, scholars design studies to test their hypotheses and further refine theories about group behavior. *Inductive* or exploratory research is also common in the social sciences, in which researchers identify a phenomenon or setting of interest and seek to better describe it or develop theory through data collection and analysis (e.g., Gersick, 1988; Ericksen & Dyer, 2004). In such cases,

researchers neither have a priori hypotheses nor are they necessarily concerned with measuring variables (see Edmondson & McManus, 2007 on research design). Because deductive research is more common, we describe it in more detail as a baseline research workflow, and only briefly summarize inductive research approaches.

Deductive research design. To design a deductive study, scholars make two related sets of decisions: (a) how to measure the variables in their hypotheses, and (b) whether to collect data from naturalistic groups (i.e., field research) or from synthetic groups formed explicitly for research purposes (i.e., laboratory research). The choice of naturalistic versus synthetic groups provides both advantages and disadvantages. Naturalistic groups exist in the field and are more realistic, thereby increasing generalizability as compared to synthetic groups. Synthetic groups are often comprised by the researcher and have no interaction history; these groups are studied in in the lab where considerably methodological control is afforded, but at the expense of real-world comparison and therefore a reduction in generalizability. Thus, the trade-off of generalizability versus control also is considered when choosing the setting for data collection.

Once researchers decide on a set of hypotheses to investigate, they must decide how to move from the theoretical concepts to actual measurement of the variables (operationalization). In group research, scholars commonly measure behavior by coding (i.e., annotating) recordings of group behaviors (see McGrath & Altermatt, 2001; Waller & Kaplan, 2016). These recordings can include video recordings, audio recordings, transcripts, or interviews about past behavior. In deductive research, coders typically do not know the hypotheses (i.e., coders are blind to hypotheses), so that their expectations do not inadvertently influence their coding of the data. Research assistants who are unaware of the studies' hypotheses often do the coding. These coders must be trained to agree with one another to a sufficient level for their codes to be

regarded as scientifically valid (i.e., inter-rater reliability). The precise level of agreement depends on the concepts, the measurements of them, and the number of coders (Hayes & Krippendorff, 2007). The coding process including inter-rater agreement is often a laborious aspect of groups research and can take weeks or months to complete depending upon amount of data to be coded.

Groups researchers can also measure members' thoughts and feelings about the group context (e.g., satisfaction, trust, cohesion) using members' self-reports, typically through survey questions (Weingart, 1997). For example, surveys of small group meetings typically contain a measure of meeting satisfaction where individuals rate a series of items (i.e., affective terms) on the degree to which they agree that they represent the recent meeting experience (see Rogelberg, Allen, Shanock, Scott, & Shuffler, 2010). Social scientists (especially social psychologists) are extremely concerned with the precise wording, sequence of questions, and response-types (e.g., 7-point agree disagree), and have many criteria for determining which to use in each situation (e.g., Sudman, Bradburn, & Schwarz, 1996). High quality survey measures have multiple questions (i.e., items) that measure the same construct/concept. Group members' responses to these questions are expected to co-vary with other items measuring the same variable (forming an index), but not with their responses to questions about other constructs on the same survey (internal validity). Valentine, Nembhart, and Edmondson (2015) provide a useful review of teamwork variables commonly measured with surveys. Further, for concepts intended to describe a group (rather than its individual members), researchers test the degree of similarity among its members' responses (within-group agreement, Hayes & Krippendorff, 2007). If data on behaviors are difficult to obtain or if the behaviors rarely occur, survey measures can be used to obtain approximate assessments of behaviors (e.g., Jehn, Rispens, & Thatcher, 2010; van der

Vegt & Bunderson, 2005; but see Baumeister, Vohs, & Funder, 2007 for concerns regarding this

behavioral proxy approach).

Deductive data collection. To test their hypotheses, social scientists need to collect data

from a sufficiently large number of groups (sample size) that is representative of their target

population of interest. Samples should contain sufficiently large numbers of groups, people, or

behaviors to have a high likelihood of detecting significant relations (i.e., statistical power*;* see

Cohen, et al., 2003)*.* The sufficiency of the sample size depends on the level of analysis; that is,

are the hypotheses about member behaviors, individual group members, or the group as a whole

(Hackman, 2002)? If researchers are collecting field data, they must obtain access to studying

groups within the organization. In some cases, this might entail video recording regular meetings

of organizational groups (e.g., Kauffeld & Lehmann-Willenbrock, 2012). In other cases, it might

require sending surveys to an organization for distribution to the relevant teams. Typically,

tailoring data collection to a specific field setting, negotiating access, and collecting group data

often takes several months. Though not particularly common, in some cases, the group

interaction data is publicly available (e.g., O'Mahony & Bechky's (2008) study of open-source

collaborations). When data on interactions already exists, it may need only to be processed and

analyzed (e.g., Stuart & Moore's (2016) study of the role of enforcers in National Hockey

League teams).

In laboratory research, researchers advertise their study and often offer individuals

compensation for participating in the study (e.g., money, course credit). Many universities

maintain *study pools,* composed of students and other members of the community willing to

participate in research. Once at the lab, researchers compose individuals into groups and give

them a task (such as a decision-making task in which information is unequally distributed among

members, e.g., Stasser & Titus, 1985; or a creativity task to make an advertisement in one hour, e.g., Gersick, 1989). The researchers then record the relevant aspects of the group interaction and measure group members' perceptions of it. Laboratory research often uses randomized-controlled experimental designs, in which groups are randomly assigned to receive either a particular *manipulation* (i.e., treatment condition) or not (control condition). For studies run with online tasks or in-class group exercises, data can be collected in a matter of hours. However, a single complex experiment can take months to conduct (e.g., Fisher's (2017) study of external interventions required the same experimenter for every group).

Deductive data analysis. In many cases, quantitative data analysis from a computer processing perspective is the fastest part of the research process. However, simply running the analysis does not include interpretation and application of the analysis. Quantitative data from all variables must be compiled into single datasets and analyzed at the appropriate level(s). Social scientists use a variety of quantitative software tools including SPSS, SAS, R, LISREL, MPLUS, HLM, and so on to analyze the data. In groups research, such analyses can be more complex than in studies of individuals. In group interactions, individual behaviors and perceptions are nested both within the group and also within the same individual over time (e.g., Krull & McKinnon, 2001). Thus, complex analyses about how and when to aggregate (or not aggregate) the data from individual behaviors to the individual level, or from individual group members to the group level, are often necessary before analyzing the data to answer the research questions or hypotheses. For variables created from coding behaviors via raters, researchers must assess whether the raters agreed on which behaviors occurred at each time (see research design section above). Inter-rater agreement on the behaviors coded must be achieved prior to any analysis of the data.

Once data is organized and coded, in many cases, researchers test whether relations among known survey items and coded data are similar to those in past, related studies. In other words, researchers seek to verify that the measures and coded variables are behaving as they should be in terms of the nature of their relationships within the data prior to further hypothesis testing. For instance, group members might be asked questions about their individual perceptions of conflict, trust, and satisfaction on the same survey. Researchers analyze whether these questions or items are associated as expected using techniques such as factor analysis (e.g., Conway & Huffcutt, 2003) and assessing the reliability of a scale (e.g., Dunn, Baguley, & Brunsden, 2014). For variables that apply to the whole group, rather than only to individual members, researchers also assess to what extent members agree with each other (within-group agreement), using statistical analyses such as the intra-class correlation coefficient and average within-group correlation (see van Mierlo, Vermunt, & Rutte, 2008). For example, after asking each member about their individual experiences of trust within the team, researchers can examine to what extent there is a climate of trust in the team (i.e., the degree to which members' perceptions of intra-team trust converge).

After statistically establishing which questions or items can be aggregated within each scale, and which scales can be aggregated within each group (or individual), researchers test their hypotheses. Groups researchers often use analytical tools based on linear regression, such as Ordinary Least-Squares, t-tests, and ANOVAs. Such analyses compute the sizes of the relations among variables and assess whether they occur substantially more often than chance (statistically significant via p-values). However, these approaches make strong assumptions about how variation is distributed within a sample and about the probabilities of finding effects. Thus, there are increasing calls to use techniques that make fewer assumptions about the underlying

distributions of variance and rely less on *p*-values to assess statistical significance (e.g., Johnson, 2013). Groups researchers are also increasingly employing analytical approaches that preserve the multi-level nature of group dynamics, rather than averaging or summing measures at the group level. These multi-level models allow researchers the freedom to specify hypotheses of the effects of groups on individual behaviors and vice versa, with greater precision (e.g., Hox, 2010).

Inductive research approaches. In inductive approaches, researchers collect exploratory data to describe new theoretical concepts, processes, and relationships. The same techniques described for deductive research above can be used to explore and describe quantitative data obtained in inductive research (Bamberger & Ang, 2016), but social scientists tend to be less inclined to ascribe causal relationship to the results of such analyses. However, there are many ways to create categorizations schemes of individual/group behavioral interaction (that can later be theorized and operationalized as variables) or show unexpected associations between variables. Such approaches are preferred when theory is nascent (i.e., there is little or no formal theory about a phenomenon; empirical description and categorization has just begun; Edmondson & McManus, 2007).

Typically, inductive researchers collect data using direct observations and interviews (Lofland, Snow, Anderson, & Lofland, 2006), although many other types of archival data are also used (e.g., press coverage; historical records; Eisenhardt, Graebner, & Sonenshein, 2016). Inductive research does NOT purport to test or prove the theories derived from the inductive methodology; that is a task for future, deductive research (Chatman & Flynn, 2005). Methods for inductive social science groups research, such as grounded theory building (Glaser & Strauss, 1967), are strongly influenced by sociology and anthropology, rather than psychology or economics. Further, inductive data typically involves detailed descriptions of phenomenon

including individual detailed accounts of the experience (Denzin, 1989). Because collecting such data is often dependent on the presence of the researcher and, as we will describe, is laborious and time-consuming to analyze, inductive researchers are typically limited to a relatively small number of cases. Inductive analyses are usually iterative—they occur concurrently with data collection, sometimes influencing what data are collected or how they are collected. The logic of such research is *bottom up*, meaning that insights that challenge the researchers' assumptions emerge from the data and affect subsequent collection and analysis.

When deciding among deductive, inductive, or mixed designs (involving both inductive and deductive aspects), groups research consider the research questions, aims, novelty of the phenomenon, and nascent nature of the domain (see Chatman & Flynn, 2005, and Edmondson & McManus, 2007 for additional details).

Computer Scientist Workflow

Computer scientists typically publish at quite a fast pace, many of these contributions are generally based on experiments completed on a computer, there can be a relatively quick turnaround to consider a hypothesis, implement it, and test it. In recent years, the pace of progress has been so great that researchers publish their work in open archives (e.g., Arxiv.org) to claim ownership of an idea rather than waiting for a conference deadline. The quality of new research papers are often determined using competitive reviewing process at yearly international conferences. The number of papers that can be presented is limited so acceptance rates at top international conferences can be much lower than for journals. Acceptance at a top international conference is often more prized than acceptance at a journal. Conference publications tend to have turnaround times of months compared to journal manuscripts that can take months to years to be finalized.

Research Design. Research design is seldom a concept that concerns computer scientists. There is no rigorous protocol to follow and more often than not, no hypotheses specified or research question expressed (see Vinciarelli et al., 2012 for exceptions and alternative perspectives). In general, a computer scientist's research question is performance driven. The prevailing, but unstated, hypothesis is that his or her proposed model is superior to all other models. Since this is the implicit starting point in this field, hypotheses are rarely spelled out. A better approach tends to break down different aspects of the model and present a more detailed understanding of where different parts fails or succeeds relative to other models. To make comparisons with the state of the art, publicly available datasets are reused hundreds of thousands of times. Using publicly available datasets is considered as showing high impact to the research community because of the open replicability of the reported results. Therefore, the data often come first, and the research questions and corresponding hypotheses are generated implicitly afterwards. Good examples of this approach include large scale efforts such as the AMI corpus, which was generated by a European-funded project with collaboration from several project partners (Carletta et al., 2005).

In describing typical workflows, computer scientists focus on social signal processing (e.g., Vinciarelli et al., 2012) and affective computing (e.g., Picard, 1997; Gunes & Schuller, 2013; Gunes & Pantic, 2010; Sariyanidi, Gunes & Cavallaro, 2015). Researchers in these two fields develop computational models to analyze behavior as accurately as possible. Behavior analysis for computer scientists has historically been about developing automated methods to interpret behavior from video. Efforts with multimodal data often combine this with audio (see surveys Gatica-Perez, 2009; Gunes & Schuller, 2013; Gunes & Pantic, 2010; Otsuka, 2011), though other forms of behavioral sensing are starting to emerge (Gunes & Hung, 2016).

Moreover, computer scientists in social signal processing are increasingly shifting away from exclusive behavior analysis to a mixture of behavior analysis and perceptions and/or cognitive processes (see surveys on meeting analysis, Gatica-Perez, 2009; behavioral sychrony, Delaherche et al., 2012; or personality estimation, Vinciarelli & Mohammadi, 2014).

Data Collection. For computer scientists, *ecological validity* refers to how closely the settings and the methods used in the computational studies replicate so-called real-world phenomena. Most past efforts focused on controlled lab-based data and settings, but recently, ecological validity has become a requirement for credible system evaluation and publication of the proposed work (Salah, Hung, Aran, & Gunes, 2013; Sariyanidi et al., 2015;). This is achieved by giving the participants the freedom to interact naturally (e.g., Salam, Celiktutan, Hupont, Gunes, & Chetouani, 2017) or by working with *found data* (i.e., using data downloaded from the web, for example, YouTube videos, Facebook images).

Sensors & Acquisition. An individual's socioemotional state may become apparent by subjective experiences (how the person feels about the other person or the situation, often assessed on a survey), by internal/inward expressions (biological signals), and by external/outward expressions (vocal/visual signals). The research problem here is to identify which of these different modalities and cues carry what communicative information to be used for continuous, fast, and efficient analysis (see Gunes & Pantic, 2010; Gunes, Piccardi, & Pantic, 2008; Gunes & Schuller, 2013 for reviews).

Annotation. Virtually all state-of-the-art computational methods for analyzing and understanding social signals and affect require labeled or annotated sets of training samples. These labels are referred to as the *ground truth*. However, there is no universal coding scheme for coding all possible individual, dyadic, or group interactions that all researchers accept and

use, and that can accommodate all possible communicative cues and modalities. Despite this, researchers obtain annotation using various tools in two ways: namely, subjective and objective evaluations. Subjective evaluations can be obtained either explicitly or implicitly. Explicit evaluations are self-report measures obtained via surveys and questionnaires (e.g., reporting the experienced emotion along the dimensions of valence, arousal, and power using the Self-Assessment Manikin (SAM); see Gunes & Schuller, 2013 for a review). Implicit evaluations can be obtained by recording participants' reactions to the displayed phenomena or to other participants (e.g., by computationally detecting their facial expressions).

Objective evaluations can be obtained by asking independent experts or observers (e.g., asking a domain expert or opt for crowd sourcing) to rate the recorded data/clip using a predefined protocol, such as the Facial Action Coding System (FACS) for Action Unit analysis (Sariyanidi et al., 2015). Labeling is done with either a discrete categorical approach (e.g., happy, sad; or personality trait categories, such as extrovert vs. introvert) or with a continuous dimensional approach for annotating the displays. Such annotations can be obtained using continuous traces in the dimensional space confined to $(-1, +1)$; e.g., arousal and valence dimensions, or the big five personality dimensions; Gunes & Schuller, 2013; Çeliktutan, Eyben, Sariyanidi, Gunes, & Schuller, 2014). When using a discrete categorical approach, if subjective labels are used the ground truth is the self-assessed label; if objective labels are used the ground truth is the mean or the median among observer ratings. Dimensional approaches consider agreement, correlation measures, and the means of across observer ratings.

Community efforts. For many years, the experimental results of different communicative behavior analysis systems could not be directly compared, which hindered progress (Sariyanidi et al., 2015). To address this issue, researchers have started to organize,

participate in, and support challenge and *benchmarking* events (e.g., Çeliktutan et al., 2014;

Schuller, 2015). Benchmarking in this context is the process of making data publicly available in

order to compare different automated analysis systems on the same data set (see Larson et al.,

2012). These events bring different communities together (e.g., researchers on speech and affect

recognition, speech and vision) and encourage them to apply and compare their approaches on a

pre-determined recognition task, under pre-defined and strictly comparable conditions.

Data analysis. As data analysis is usually built on machine-interpreted strategies, there

are no strong *a priori* assumptions on the mechanisms that generated the data, unlike most

statistics. Social signal processing and affective computing employ pattern recognition to

develop machines that can observe data (e.g., objects, signals, images), identify distinguishable

patterns of interest, and make decisions about the categories of patterns (Duda, Hart, & Stork,

2001). For computer scientists, a pattern denotes a vector of numerical values that describe data.

To obtain a p-dimensional vector, we measure p-features on data. For example, we consider that

a good representation of an image could be the intensity of each pixel. Pattern recognition

processing techniques, such as data comparison or matching, require a feature vector based

representation.

To classify data reliably, pattern recognition systems require that observed behavioral

data undergo several processing steps: pre-processing, feature selection and/or extraction,

classification of these features into categories called *classes* (e.g., discrete emotion), or

prediction of a numerical value (e.g., valence of an emotion).

Pre-processing. Sensor data collection usually generates *heterogeneous* signals with

noise, missing and/or out-of-range values, which are errors usually called artifacts. Data pre-

processing reduces these artifacts and noise, and often includes mean and variance normalization

of data (i.e., translating data into z-scores to make data across collection contexts comparable).

The issue is how to determine the relevant normalization schemes. For example, given a small

group of people, if the research question addresses individual or group differences, computer

scientists rely on adequate normalization schemes: individual level normalization or group level

normalization of signals.

      Feature selection and extraction. *Features* are measurements on the data specified by

the researchers. For example, to recognize speakers' intonation styles, computer scientists would

leverage expertise in prosody linguistics-related components, such as analyzing the fundamental

frequency of the speech signal. There is a growing need to improve the interpretation of such

features in various domains such as clinical settings (Leclère et al., 2016), organizational

psychology (Nguyen, Frauendorfer, Schmid-Mast, Gatica-Perez, 2014) and human-machine

interaction (Salam, Celiktutan, Hupont, Gunes, & Chetouani, 2017). The challenge here is

generate features that could be interpreted by humans. For example, to model human's

engagement during interaction, features such as head and body orientation are interpretable by

researchers of various domains.

      Feature selection is a process that reduces the number of features by selecting a subset of

the original features. Features that are not considered discriminant enough are then simply

removed from the feature vector. *Feature extraction* is a process that reduces the number of

features by transforming the original full set of available features into a small subset of features.

The projection of the original feature vectors to a lower dimensional space is a common

transformation technique. The transformation can be either a linear or nonlinear combination of

features, so interpreting the extracted features is not always straightforward. For example,

Principal Component Analysis (PCA) is a linear feature extraction technique that transforms the

original features into a set of linearly uncorrelated features. In computer vision, PCA is employed to generate *eigenfaces*, which are obtained by transforming an original set of images into a smaller set of basis images. This low-dimensional representation can be linearly combined to reconstruct the original set of images. An extension to speech with eigenvoices and behaviors with eigenbehaviors are also possible.

Whether feature extraction belongs to the classification process or not is an old debate in the pattern recognition community. Renewing this debate, recently so-called *end-to-end models* have been proposed to jointly optimize feature extraction and classification (Graves & Jaitly, 2014). End-to-end models have been developed to decrease the dependency on human a priori knowledge in the design of the feature extraction step by learning an adapted representation to the subsequent classification step. This approach is new trend in the computer science community.

Decision, classification, and prediction. Making a decision, either using classification or prediction, is essential for the machine-based approach. Using a classification approach provides automatic *categorization* of the patterns into classes following the discrete categorical approach (e.g., six basic emotions). When using a prediction approach, estimating continuous dimensions (e.g., valence), regression or other predictive models are used.

Classification separates patterns into classes using mathematical functions and machine-learning techniques provide an automatic way to optimize it. For example, the minimum-distance classifier estimates a single representative by class (usually the mean of all feature vectors of the class). Then, the classification is simply performed by assigning the unlabeled patterns to the nearest representative. The mathematical properties (e.g., linear, polynomial, recurrent) of the functions highly impact the performance of the classifier and they are usually set-up by the

researcher. However, this step is complex, as oversimplified models (e.g., free of parameters) omit key aspects of the data's key structures (underfitting), while excessively complex models capture irrelevant structures in the data (overfitting). For example, setting very low or high number of representatives for each class necessarily impacts the performance and is empirically performed.

Evaluation. The appropriate manner to evaluate the quality of a computational model is to optimize it using *training data* and evaluate the generalization capabilities using *testing data*. For example, regarding the minimum-classifier distance, training data will be employed to estimate the representatives of each class (e.g., mean of each class). Testing data will be employed to estimate the performance of the classifier using unseen data. Partitioning data into learning and testing subsets is challenging. *Cross-validation* approaches, which allow distinct partitions of training and testing data are preferred. Natural and common metrics to evaluate classification and prediction tasks include error rate (e.g., accuracy) and *mean squared error (MSE)*. For example, the accuracy measures the fraction of all feature vectors that are correctly categorized by the minimum-classifier distance.

Similarities and Differences between Social and Computer Scientists' Workflows

Through a review of the previous sections, several similarities and poignant differences emerge. As the previous sections were not intended to be comprehensive, neither is this section intended to be exhaustive in its treatment of similarities and differences in the workflows of social and computer scientists. Instead, the purpose of this section is to simply share some of the similarities and differences that likely exist and point to how these similarities and complimentary differences can be leveraged toward a win-win approach. It must also be noted that our identification of similarities and differences is heuristic and may not always hold. That

is, we suggest that many social and computer scientists share these similarities and many share these differences. However, not all social scientists and computer scientists will necessarily fit within these generalizations, and there will certainly be exceptions on both sides. The purpose is to capture some of the tendencies and trends, corroborate some of ours and others' experiences, and help plot a way forward for collaboration at the highest level.

The workflows of both social and computer scientists have important similarities that should enable collaboration. First, both share the goal of modeling human behavior, seeking to understand why people do what they do, and how they interact with each other and their environments using rigorous and replicable methods. Second, both workflows often rely on video recordings of social interactions that are coded/annotated by independent raters, with the goal of obtaining reliable and objective observations. Third, both social and computer scientists often use interactants' self-reports of their internal states via surveys. Fourth, both have some commonality in the analytic tools which they typically use, such as factor analysis and measures of agreement, albeit with different purposes in mind. Additionally, both disciplines face the challenge of how to synthesize the results of different studies to obtain generalizable knowledge. In other words, social scientists and computer scientists share a remarkable amount of similarities as they both seek to understand human behavior. These similarities can help them to overcome or integrate the differences that make interdisciplinary collaboration challenging.

In other ways, the workflows of social and computer scientists have important differences in research design, data detail and validity, analyses, and publications. Whereas social scientists spend more effort on research design and data collection, computer scientists often spend little time on data collection or research design and instead focus their efforts on iterative processes of analysis. Whereas social scientists emphasize hypothesized associations between different

phenomena or variables (often with a concern for group performance), computer scientists emphasize the efficiency of predictive models in matching annotations and self-reports.

The detail of the data and concerns about their validity also differ across fields. Computer scientists tend to use more different types of sensing modalities to record human behavior than social scientists, who have focused mostly on audio or video recordings. As a result, computer scientists focus on fine-grained aspects of social behavior (e.g., facial expressions, behavioral mimicry), while social scientists tend to focus more on more general aspects of interaction (e.g., complaining cycles, socioemotional interaction). As such, social scientists will use audio or video recordings whereas a computer scientist may want additional, more nuanced recording tools such as accelerometers to capture additional detail from the interaction. While social scientists emphasize inter-rater agreement and internal validity for survey and observational measures (i.e., whether they actually measuring our target construct), computer scientists are less concerned with this, and instead embrace some subjectivity in the data as capturing the real messy and subjective world.

Importantly, the types of publications and timelines differ widely between the fields, which can create conflicting incentives, especially for early career researchers. Social and computer scientists tend to publish in different venues and the time to publication varies dramatically. For a computer scientist, one completed paper could easily have been the result of 50 to 100 iterations of different ideas tested on the same data. Social scientists on the other hand, usually collect unique data for each set of hypotheses, often including multiple datasets in the same paper. Reusing data for the same analyses is considered unethical by most social science journals. Hence, negotiating between these differences creates one of the largest practical obstacles to synergistic collaboration.

As these similarities and differences illustrate, collaborations between computer scientists and social scientists offer potentially fruitful opportunities but will require understanding and patience. The similarities can facilitate mutual understanding, and the differences may require patience as the ambiguity created by inexperience with each other's workflow can create questions or constructive conflict.

Towards a Win-Win Approach

Despite the challenges presented by differing workflows, we believe that social science and computer science researchers can make inroads toward long-lasting and synergistic interdisciplinary collaborations. These collaborations can involve differing types of interdependence between researchers for the two disciplines (see Figure 1). The producer-consumer model to which we referred earlier involves *pooled interdependence* (Thompson, 1967). In such a setup, there is little direct interaction between disciplines, but rather each discipline relies on the other to produce adequate resources for the other. This means that no collaboration between researchers occurs; producers and consumers publish independently of one another and benefit from each other only via publications or other publicly shared resources (e.g., Hung & Gatica-Perez, 2010; Jayagopi, Hung, Yeo, & Gatica-Perez, 2009).

Although the producer-consumer model is not considered collaboration, it is an important basis for considering where entry points of collaboration could occur. Such an approach enables researchers to identify potential interdisciplinary partners via common research questions or data. Finding out about the tools and practices in the other disciplines and implementing them in one's own research is a critical first step toward more complex collaborations. We argue that social scientists should use available computer science tools when possible, and that computer scientists should seek out social scientists producing annotated videos. Further, social scientists

should use recording equipment of a higher quality which makes it possible for computer

scientists to use the recorded data. To facilitate this, computer scientists can aggregate and list

what tools are available, and what dimensions can be automatically coded.

Interdisciplinary collaboration can also occur when a paper written for one discipline has

minor contributions in the other discipline. A scholar from the other discipline may then become

a co-author (e.g., Sanchez-Cortes, Aran, Jayagopi, Schmid-Mast, & Gatica-Perez, 2013). This

implies *sequential interdependence,* in which work is analogous to an assembly line where each

person contributes and hands off the product to the next (Thompson, 1967). In this model, social

scientists and computer scientists each contribute to the same project, each fulfilling discrete

tasks, and then passing off to the other group. For instance, social scientists collecting new data

on meetings in organizations could annotate video recordings for behaviors of interest to

computer scientists (while also annotating for other behaviors or states specific to their own

research). Those annotated videos could be passed off to the computer scientists, who then use

them for training and evaluation, but might also help the social scientists implement automated

coding of emotions from facial expressions or vocal prosody for use in their project. In such an

arrangement, each discipline benefits from the other and circumvents potential problems related

to the different pace at which both disciplines go from research to publication. This model is

open to a broad range of computer scientists and social scientists who may not be familiar with

interdisciplinary collaboration.

At the next level of collaboration, joint research questions are formed in which the

producer-user roles are more heavily interwoven with resources provided by the producer. For

instance, a social scientist could take the lead in providing research questions and hypotheses and

code for behaviors, whereas the computer scientist would provide novel automated approaches to

test or verify the hypotheses. The data analysis step would therefore use the expertise of both

computer scientists and and social scientists. A similar level of integration could be reached

when a common data set is collected, but both disciplines publish independently from the data as

well as together.

The highest level of interdisciplinary collaboration—*reciprocal interdependence*—is the

most challenging to obtain. Reciprocal interdependence involves the actions of both parties being

tightly interwoven and requires constant mutual adjustment, as in a basketball team or jazz

ensemble (Thompson, 1967). This is likely to be best undertaken by senior researchers who

already have some experience working in pooled or sequential interdependence arrangements.

Because neither field is organized to accommodate the other (e.g., in terms of journals,

promotion/tenure procedures), such interdisciplinarity involves large risks. Specifically, what

counts as a publication in a reputable outlet and therefore supportive a tenure portfolio is

different for each discipline and potentially unique to each department/university. However,

seeking opportunities after undertaking the looser forms of interdependence, and developing

research questions and approaches through these processes will eventually result in outcomes

that have mutual benefit (e.g., each publishing in their respective disciplines' journals). The

expectation here, and one of the initial motivations for the Lorentz workshop and for this

resulting special issue, is that the mutual benefits at this level will be so sufficiently ground-

breaking for each discipline in its own right that the time investment will be worth it. Essentially,

starting the conversation about the opportunities for truly collaborative interdisciplinary research

will bear fruit soon.

It is our hope that as computer scientists and social scientists come together and start to

observe the high gain that both sides can accomplish, this higher level of mutual understanding

will become the norm of an integrated community rather than pockets of interdisciplinary

collaborations. Through our review of the workflows, several barriers to and synergies towards

collaboration were identified. The identification of these barriers and synergies is a necessary

first step for the mutually beneficial collaboration described here. Specifically, as collaborations

form, we hope that researchers will invest the time, energy, and resources necessary to arrive at

the highest level of integrated collaboration in a true interdisciplinary and ground breaking

fashion.

References

Bamberg, P. & Ang, S. (2016). The quantitative discovery: What it is and how to get it

published. *Academy of Management Discoveries, 2*, 1-6. doi:10.5465/amd.2015.0060

Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-

reports and finger movements: Whatever happened to actual behavior? *Perspectives on*

Psychological Science, *2*, 396–403. doi:10.1111/j.1745-6916.2007.00051.x

Carletta, J., Ashby, S., Bourban, S., Flynn, M., Guillemot, M., Hain, T., . . . Lathoud, G. (2005,

July). The AMI meeting corpus: A pre-announcement. In S. Renals & S. Bengio (Eds.),

International Workshop on Machine Learning for Multimodal Interaction (pp. 28-39).

doi:10.1007/11677482_3

Çeliktutan, O., Eyben, F., Sariyanidi, E., Gunes, H., Schuller, B. W. (2014). MAPTRAITS 2014:

The first audio/visual mapping personality traits challenge--an introduction: Perceived

personality and social dimensions. *Proceedings of the International Conference on*

Multimodal Interaction, Turkey, 529-530.

Chatman, J. A., & Flynn, F. J. (2005). Full-cycle micro-organizational behavior research.

Organization Science, *16*, 434–447. doi:10.1287/orsc.1050.0136

Conway, J. M., & Huffcutt, A. I. (2003). A review and evaluation of exploratory factor analysis

practices in organizational research. *Organizational Research Methods, 6*, 147–168.

doi:10.1177/1094428103251541

Delaherche, E., Chetouani, M., Mahdhaoui, A., Saint-Georges, C., Viaux, S., & Cohen, D.

(2012). Interpersonal synchrony: A survey of evaluation methods across disciplines. *IEEE*

Transactions on Affective Computing, 3, 349-365. doi:10.1109/t-affc.2012.12

Denzin, N. K. (1989). *Interpretive biography*. Newbury Park, CA: Sage.

Duda, R. O., Hart, P. E., Sork, D. G. (2001) *Pattern classification* (2nd ed.). New York, NY: Wiley.

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, *105*, 399–412. doi:10.1111/bjop.12046

Edmondson, A. C., & McManus, S. E. (2007). Methodological fit in management field research. *Academy of Management Review*, *32*, 1155–1179. doi:10.5465/AMR.2007.26586086

Eisenhardt, K. M., Graebner, M. E., & Sonenshein, S. (2016). Grand challenges and inductive methods: Rigor without rigor mortis. *Academy of Management Journal*, *59*, 1113–1123. doi:10.5465/amj.2016.4004

Ericksen, J., & Dyer, L. (2004). Right from the start: Exploring the effects of early team events on subsequent project team development and performance. *Administrative Science Quarterly*, *49*, 438–471. doi:10.2307/4131442

Fisher, C. M. (2017). An ounce of prevention or a pound of cure? Two experiments on in-process interventions in decision-making groups. *Organizational Behavior and Human Decision Processes*, *138*, 59–73. doi:10.1016/j.obhdp.2016.11.004

Gatica-Perez, D. (2009). Automatic nonverbal analysis of social interaction in small groups: A review. *Image and Vision Computing, 27*, 1775-1787.

Gersick, C. J. G. (1988). Time and transition in work teams: Toward a new model of group development. *Academy of Management Journal*, *31*, 9–41. doi:10.2307/256496

Gersick, C. J. G. (1989). Marking time: Predictable transitions in task groups. *Academy of Management Journal*, *32*, 274–309. doi: 10.2307/256363

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago, IL: Aldine Transaction.

Graves, A., Jaitly, N. (2014). Towards end-to-end speech recognition with recurrent neural

networks. *Proceedings of the 31st International Conference on Machine Learning, PMLR

32*, 1764-1772.

Gunes, H. & Pantic, M. (2010). Automatic, dimensional and continuous emotion recognition.

International Journal of Synthetic Emotions, 1(1), 68-99. doi:10.4018/jse.2010101605

Gunes, H. & Schuller, B. (2013). Categorical and dimensional affect analysis in continuous

input: Current trends and future directions. *Image and Vision Computing*, *31*, 120-136.

doi:10.1016/j.imavis.2012.06.016

Gunes, H., & Hung, H. (2016). Is automatic facial expression recognition of emotions coming to

a dead end? The rise of the new kids on the block. *Image and Vision Computing, 55*, 6-8.

doi:10.1016/j.imavis.2016.03.013

Gunes, H., Piccardi, M. & Pantic, M. (2008). From the lab to the real world: Affect recognition

using multiple cues and modalities. In J. Or (Ed.), *Affective computing, focus on emotion

expression, synthesis and recognition* (pp. 185-218). Vienna, Austria: I-Tech Education and

Publishing.

Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure

for coding data. *Communication Methods and Measures*, *1*, 77–89.

doi:10.1080/19312450709336664

Hox, J. J. (2010). *Multilevel analysis: Techniques and applications* (2nd ed.). New York:

Routledge.

Hung, H., & Gatica-Perez, D. (2010). Estimating cohesion in small groups using audio-visual

nonverbal behavior. *IEEE Transactions on Multimedia*, *12*, 563-575.

doi:10.1109/TMM.2010.2055233

Jayagopi, D. B., Hung, H., Yeo, C., & Gatica-Perez, D. (2009). Modeling dominance in group

conversations using nonverbal activity cues. *IEEE Transactions on Audio, Speech, and*

Language Processing, *17*, 501-513. doi:10.1109/TASL.2008.2008238

Jehn, K. A., Rispens, S., & Thatcher, S. M. B. (2010). The effects of conflict asymmetry on work

group and individual outcomes. *Academy of Management Journal*, *53*, 596–616.

doi:10.5465/AMJ.2010.51468978

Johnson, V. E. (2013). Revised standards for statistical evidence. *Proceedings of the National*

Academy of Sciences, *USA, 110*, 19313–19317. doi:10.1073/pnas.1313476110

Kauffeld, S., & Lehmann-Willenbrock, N. (2012). Meetings matter: Effects of team meetings on

team and organizational success. *Small Group Research*, *43*, 130–158.

doi:10.1177/1046496411429599

Krull, J. L., & MacKinnon, D. P. (2001). Multilevel modeling of individual and group level

mediated effects. *Multivariate Behavioral Research*, *36*, 249-277.

doi:10.1207/s15327906mbr3602_06

Larson, M., Soleymani, M., Eskevich, M., Serdyukov, P., Ordelman, R., & Jones, G. (2012). The

Community and the Crowd: Multimedia Benchmark Dataset Development. *IEEE*

MultiMedia, *19*(3), 15-23. doi:10.1109/MMUL.2012.27

Leclère C., Avril M., Viaux-Savelon S., Bodeau N., Achard C., Missonnier S., . . . Cohen D.

(2016), Interaction and behaviour imaging: A novel method to measure mother-infant

interaction using video 3D reconstruction. *Translational Psychiatry 6*, e816 ; doi:

10.1038/tp.2016.82

Lofland, J., Snow, D. A., Anderson, L., & Lofland, L. H. (2006). *Analyzing social settings*.

Belmont, CA: Wadsworth Publishing Company.

McGrath, J. E., & Altermatt, T. W. (2001). Observation and analysis of group interaction over

 time: Some methodological and strategic choices. In M. A. Hogg & R. S. Tindale (Eds.),

 Blackwell handbook of social psychology: Group processes (pp. 525–556). Malden, MA:

 Blackwell.

Mohr, L. B. (1982). *Explaining organizational behavior*. San Francisco, CA: Jossey-Bass.

Nguyen, L. S., Frauendorfer, D., Mast, M. S., Gatica-Perez, D (2014). Hire me: Computational

 inference of hirability in employment interviews based on nonverbal behavior. *IEEE*

 Transactions on Multimedia, *16*, 1018-1031. doi:10.1109/TMM.2014.2307169

O'Mahony, S., & Bechky, B. A. (2008). Boundary organizations : Enabling collaboration among

 unexpected allies. *Administrative Science Quarterly , 53,* 422-459.

 doi:10.2189/asqu.53.3.422

Otsuka, K. (2011). Conversation scene analysis [social sciences]. *IEEE Signal Processing*

 Magazine, *28*(4), 127-131. doi:10.1109/MSP.2011.941100

Picard, R. W. (1997). *Affective computing*. Cambridge, MA: MIT Press

Rogelberg, S. G., Allen, J. A., Conway, J., Goh, A., Currie, L., & McFarland, B. (2010).

 Employee experiences with volunteers: Assessment, description, antecedents, and outcomes.

 Non-Profit Management and Leadership Journal, 20, 423-444. doi:10.1002/nml.20003

Salah, A. A., Hung, H., Aran, O. & Gunes, H. (2013). Creative applications of human behavior

 understanding. In A. A. Salah, H. Hung, O. Aran, & H. Gunes (Eds.), *Lecture notes in*

 computer science, 8212. Human Behavior Understanding (pp. 1-14). doi:10.1007/978-3-

 319-02714-2_1

Salam H., Celiktutan O., Hupont I., Gunes H., & Chetouani M. (2017), Fully automatic analysis

    of engagement and its relationship to personality in human-robot interactions. *IEEE Access*,

    5, 705-721. doi:10.1109/ACCESS.2016.2614525

Sanchez-Cortes, D., Aran, O., Jayagopi, D. B., Mast, M. S., & Gatica-Perez, D. (2013).

    Emergent leaders through looking and speaking: From audio-visual data to multimodal

    recognition. *Journal on Multimodal User Interfaces*, *7*(1-2), 39-53. doi:10.1007/s12193-

    012-0101-0

Sariyanidi, E., Gunes, H., & Cavallaro A. (2015). Automatic analysis of facial affect: A survey

    of registration, representation, and recognition. *IEEE Transactions onPattern Analysis and

    Machine Intellingence* 37, 1113-1133. doi:10.1109/TPAMI.2014.2366127

Schuller, B. (2015) Multimodal Ddatabases: Collection, challenges, and chances. In R. A. Calvo,

    S. D'Mello, J. Gratch, & A. Kappas (Eds.), *The Oxford handbook of affective computing*

    (pp. 323-333). New York, NY: Oxford University Press.

Stasser, G., & Titus, W. (1985). Pooling of unshared information in group decision making: Biased

    information sampling during discussion. *Journal of Personality and Social Psychology*, *48*, 1467–

    1478. doi:10.1037/0022-3514.48.6.1467

Stuart, H. C., & Moore, C. (2016). Shady characters: The implications of illicit organizational roles for

    resilient team performance. *Academy of Management Journal*. Advance online publication.

    doi:10.5465/amj.2014.0512

Sudman, S., Bradburn, N. M., & Schwarz, N. (1996). *Thinking about answers: The application

    of cognitive processes to survey methodology.* San Francisco, CA: Jossey-Bass.

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts

    and applications*. Washington, DC: American Psychological Association.

Thompson, J. D. (1967). *Organizations in action: Social science bases of administrative theory*.

New Brunswick, NJ: Transaction.

Valentine, M. A., Nembhard, I. M., & Edmondson, A. C. (2015). Measuring teamwork in health

care settings: A review of survey instruments. *Medical Care*, *53*(4), 16–30.

doi:10.1097/MLR.0b013e31827feef6

van der Vegt, G. S., & Bunderson, J. S. (2005). Learning and performance in multidisciplinary

teams: The importance of collective team identification. *Academy of Management Journal*,

48, 532–547. doi:10.5465/AMJ.2005.17407918

van Mierlo, H., Vermunt, J. K., & Rutte, C. G. (2008). Composing group-level constructs from

individual-level survey data. *Organizational Research Methods*, *12*, 368–392.

doi:10.1177/1094428107309322

Vinciarelli, A., & Mohammadi, G. (2014). A survey of personality computing. *IEEE*

Transactions on Affective Computing, 5, 273-291. doi:10.1109/TAFFC.2014.2330816

Vinciarelli, A., Pantic, M., Heylen, D., Pelachaud, C., Poggi, I., D'Errico, F., & Schroeder, M.

(2012). Bridging the gap between social animal and unsocial machine: A survey of social

signal processing. *IEEE Transactions on Affective Computing*, *3*, 69-87. doi:10.1109/T-

AFFC.2011.27

Wageman, R. (2001). How leaders foster self-managing team effectiveness: Design choices

versus hands-on coaching. *Organization Science*, *12*, 559–577.

doi:10.1287/orsc.12.5.559.10097

Wagner, C. S., Roessner, J. D., Bobb, K., Klein, J. T., Boyack, K. W., Keyton, J., . . . Börner, K.

(2011). Approaches to understanding and measuring interdisciplinary scientific research

(IDR): A review of the literature *Journal of Informetrics*, *5*, 14-26.

doi:10.1016/j.joi.2010.06.004

Waller, M. J., & Kaplan, S. A. (2016). Systematic behavioral observation for emergent team

pPhenomena: Key considerations for quantitative video-based approaches. *Organizational*

*Research Methods*. Advance online publication. doi:10.1177/1094428116647785

Weingart, L. R. (1997). How did they do that? The ways and means of studying group process.

Research in Organizational Behavior, *19*, 189–239.

Table 1: Summary of Workflows for Social and Computer Scientists

Workflow Domain	Social Scientist	Computer Scientist
1. Research Design	• Deductive or inductive designs typically using quantitative/qualitative data with the intent of test or developing theory	• Typically performance-driven; the author's proposed model is better than anyone else's.
2. Data Collection	• Gather data from adequate number of groups using survey tools, audio/video recording equipment, or some combination of collection tools	• Focuses on acquisition of sensor data of human behavior. Typically, data and associated annotations are shared and reused in the research community.
3. Data Analysis	• Typically involves quantitative analysis via computer software. • Often includes coding of behaviors recorded using audio/video files. • Inter-rater agreement on coded behavior is expected. • Takes the form of hypothesis test for deductive designs or theory building for inductive designs.	• Typically, automated methods to estimate the label of sensor data are developed. • The proposed methods are always tested using different data than the data used for training the model. • Performance is expected to be validated using cross-validation.

Figure 1: Interdisciplinary Collaboration via Interdependence Level and Time

Early
Collaborative
Interdependence

Pooled   Sequential   Reciprocal

Later
Collaborative
Interdependence

TIME
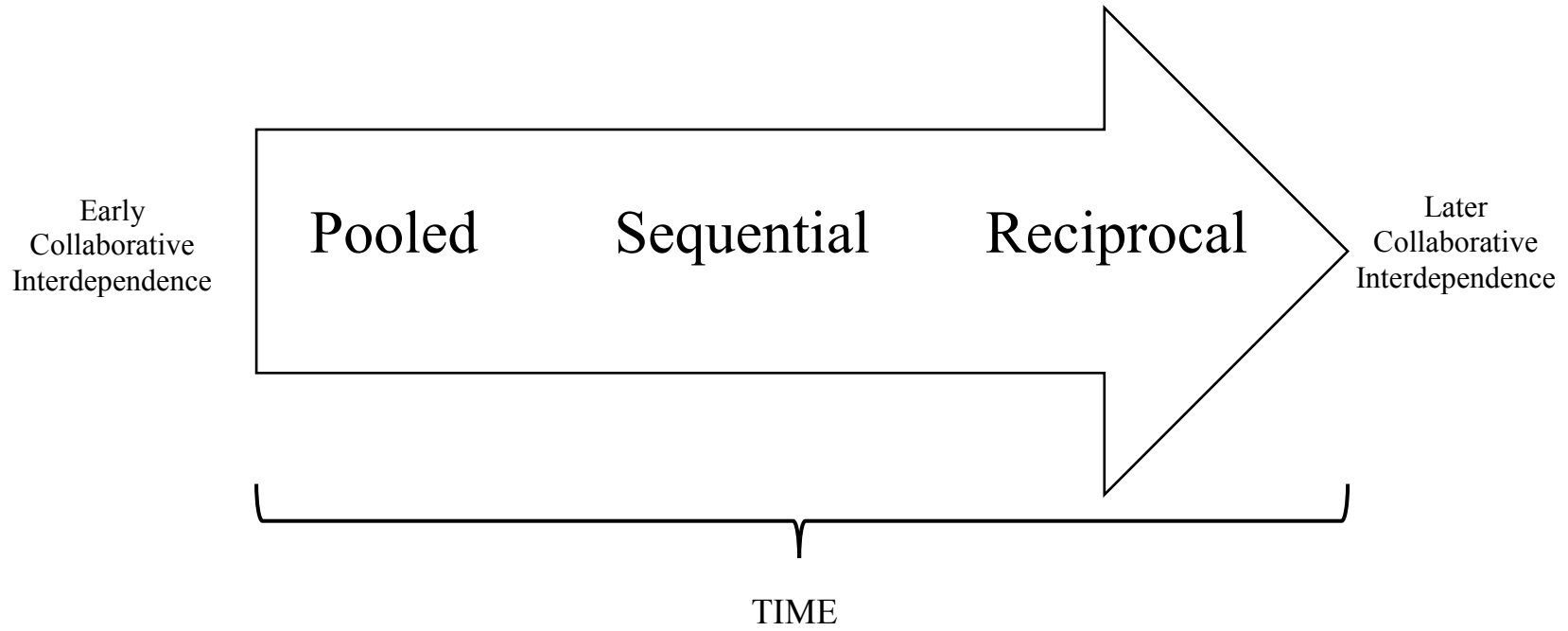
Author Bios

Joseph A. Allen is an Associate Professor of Industrial and Organizational Psychology at the University of Nebraska at Omaha, USA. His research deals with meeting science, the management of volunteers in nonprofit organizations, and emotional labor.

Colin M. Fisher is an Assistant Professor of organizational behavior at University College London, England. His research deals with collaboration and temporal dynamics (i.e., timing, rhythm, development over time) in three areas: leading, helping, and coaching teams; collective creativity and improvisation; and group decision making and negotiations.

Mohamed Chetouani is a full professor in signal processing and machine learning for human-machine interactions at the Institute of Intelligent Systems and Robotics, Sorbonne University, University Pierre and Marie Curie, France. His research activities cover the areas of social signal processing and personal robotics.

Ming Ming Chiu is Charles R. Hicks Professor in the College of Education at Purdue University, USA. He invented statistical discourse analysis (SDA), multilevel diffusion analysis (MDA), and an artificial intelligence expert system. He is developing methods for automatic computer analysis of online discussions and detection of online sexual predators.

Hatice Gunes is an Associate Professor in the Computer Laboratory, Faculty of Science and Technology, University of Cambridge, England. Her research expertise is in affective computing and social signal processing across multiple disciplines including computer vision, signal processing, machine learning, multimodal interaction and human–robot interaction.

Marc Mehu is Assistant Professor of Psychology at Webster Vienna Private University, Austria. His research interests lie in the study of nonverbal communication and emotional expressions from social psychological and evolutionary perspectives.

Hayley Hung (PhD., Queen Mary University of London), is an Assistant Professor at Delft University of Technology, The Netherlands. She is an expert in social signal processing. Her research contributions have recently been recognized internationally during the Rising Stars Symposium of the Association for Computing Machinery Multimedia conference in 2016.