

## An Application of Extreme Value Theory to Learning Analytics: Predicting Collaboration Outcome from Eye-tracking Data

**Kshitij Sharma**

Department of Operations, Faculty of Business and Economics  
University of Lausanne, Switzerland  
Computer Human Interaction in Learning and Instruction  
School of Computer and Communication Sciences  
École Polytechnique Fédérale de Lausanne, Switzerland  
[kshitij.kshitij@unil.ch](mailto:kshitij.kshitij@unil.ch)

**Valérie Chavez-Demoulin**

Department of Operations, Faculty of Business and Economics  
University of Lausanne, Switzerland

**Pierre Dillenbourg**

Computer Human Interaction in Learning and Instruction  
School of Computer and Communication Sciences  
École Polytechnique Fédérale de Lausanne, Switzerland

**ABSTRACT:** The statistics used in education research are based on central trends such as the mean or standard deviation, discarding outliers. This paper adopts another viewpoint that has emerged in statistics, called extreme value theory (EVT). EVT claims that the bulk of normal distribution is comprised mainly of uninteresting variations while the most extreme values convey more information. We apply EVT to eye-tracking data collected during online collaborative problem solving with the aim of predicting the quality of collaboration. We compare our previous approach, based on central trends, with an EVT approach focused on extreme episodes of collaboration. The latter provided a better prediction of the quality of collaboration.

**KEYWORDS:** Eye-tracking, dual eye-tracking, extreme value theory, computer supported collaborative learning, learning analytics, collaboration quality

## 1 INTRODUCTION

This contribution borrows a framework from the field of statistics called extreme value theory (EVT), which has been developed for analyzing time series in domains such as finance and environmental sciences. We explore the relevance of EVT for learning analytics, namely for analyzing collaborative interactions in an educational setting. For these kinds of analyses, statistical methods traditionally focus

(2017). An application of extreme value theory to learning analytics: Predicting collaboration outcome from eye-tracking data. *Journal of Learning Analytics*, 4(3), 140–164. <http://dx.doi.org/10.18608/jla.2017.43.8>

on the central tendencies (mean, median, and standard deviation). Generally, we discarded what we considered to be outliers, which we suspected might be due to measurement errors, cheating, or miscellaneous events foreign to the cognitive mechanisms under scrutiny. Instead, EVT invites us to focus on the interaction episodes, which deviate from those central tendencies. The shift between these two approaches, from central to extremes, is accompanied by another shift: the extreme data points do not correspond to an individual subject or a pair but to some specific time episodes within a long series of time events produced by each individual or pair. The goal of this paper is to determine if EVT could provide us with better discrimination among different levels of collaboration quality compared to traditional methods. We therefore apply both methods to the times series produced by eye trackers and compare the results. Since we study collaboration, we synchronized the eye-tracking data produced by each peer (what we call “dual eye-tracking”). EVT has been traditionally used to quantify rare events like century floods, avalanches, market crashes, or more recently terrorism attacks. Outside of the risk management context, it has not been much developed because of the lack of rare data. In this paper, we propose the use and development of extreme value learning tools to explore “rare data” from educational “big data” experiments such as eye-tracking experiments.

The paper is organized as follows: Section 2 describes the nature of dual eye-tracking data (DUET), followed in Section 3 by an introduction to EVT. Section 4 introduces the concept that bridges DUET and EVT in two ways. In the univariate way, each pair of time episodes from learners A and B is substituted by a measure of their differences, which produces a time series of single values. In the bivariate mode, we take into consideration the dynamic coupling of the two time series. The rest of the paper compares the results produced by EVT to those resulting from traditional approaches.

## 2 EYE-TRACKING

Eye-tracking provides researchers with unprecedented access to information about users’ attention. The eye-tracking data is rich in terms of temporal resolution. With the advent of eye-tracking technology, the eye-tracking apparatus has become compact and easy to use without sacrificing much of its ecological validity during the controlled experiments. Previous research had shown that eye-tracking can be useful for unveiling the cognitive processes that underlie verbal interaction and problem-solving strategies. We introduce here some key concepts necessary to understand the study presented later.

### 2.1 Fixations and Saccades

In a nutshell, gaze does not glide over visual material in a smooth continuous way but rather jumps around the stimulus: small stops around 200 milliseconds, called “fixations,” are followed by long jumps, called “saccades.” It is hypothesized that information is collected only during fixations. However, the data analysis is more complex. What if the eyes stop after 180 or 170 milliseconds? Can this still be considered as a fixation? Eye-tracking methods require different thresholds to be defined in order to process data. Are these thresholds the same for all subjects and for all tasks? If we consider a single subject on a single task, is the threshold stable over time? Is it the same in the middle of the screen or on the periphery? Eye-tracking relies on the craft of “thresholding.” Nüssli (2011) developed

(2017). An application of extreme value theory to learning analytics: Predicting collaboration outcome from eye-tracking data. *Journal of Learning Analytics*, 4(3), 140–164. <http://dx.doi.org/10.18608/jla.2017.43.8>

optimization algorithms that systematically explore threshold parameters in order to maximize the quality of produced data. Several studies have shown that the level of expertise of an individual (Ripoll, Kerlirzin, Stein, & Reine, 1995; Abernethy & Russell, 1987; Charness, Reingold, Pomplun, & Stampe, 2001; Reingold, Charness, Pomplun, & Stampe, 2001) could be determined from eye-tracking data since the way one looks at an X-RAY (Grant & Spivey, 2003; Thomas & Lleras, 2007) or a piece of programming code (Sharma, Jermann, Nüssli, & Dillenbourg, 2012) reveals the way one understands these things. We will not develop these findings in this paper as we focus on collaborative situations. For instance, within a collaborative Tetris game, Jermann, Nüssli, and Li (2010) predicted the level of expertise in a pair (expert–expert, novice–novice, or expert–novice pair) with an accuracy of 75%. The core relationship between gaze and collaboration results from the gaze-dialogue coupling.

## 2.2 Gaze-dialogue Coupling

Two eye-trackers can be synchronized for studying the gaze of two persons interacting to solve a problem and for understanding how gaze and speech are coupled. Meyer, Sleiderink, and Levelt (1998) showed that the duration between looking at an object and naming it is between 430 and 510 milliseconds (eye–voice span). Griffin and Bock (2000) found an eye–voice span of about 900 milliseconds. Zelinsky and Murphy (2000) discovered a correlation between the time spent gazing at an object and the spoken duration the name of the object was given aloud. Richardson, Dale, and Kirkham (2007) proposed the eye–eye span as the difference between the time when the speaker starts looking at the referred object and the time when listeners look at it. This time lag was termed the “cross-recurrence” between the participants. The average cross-recurrence was found to be between 1,200 and 1,400 milliseconds. Jermann and Nüssli (2012) applied cross-recurrence to a pair programming task, enabling the remote collaborators to see their actions on the screen. The authors found that the cross-recurrence levels were higher when selection was mutually visible on the screen, which related to the cross-recurrence of team coordination.

## 2.3 Quality of Interaction and Cross-recurrence

Several authors have found a relationship between the cross-recurrence of gazes and the quality of collaboration. Cherubini and Dillenbourg (2007) found a correlation between gaze-recurrence and the performance of teams in a map annotation task. In a peer programming task, Jermann and Nüssli (2012) found higher gaze recurrence for pairs that collaborate well, as estimated by the Meier, Spada, and Rummel (2007) qualitative coding scheme. In a concept-map task (Sharma, Caballero, Verma, Jermann, & Dillenbourg, 2015; Sharma, Jermann, Nüssli, & Dillenbourg, 2013) related cross-recurrence to higher learning gains. In a collaborative learning task using tangible objects, Schneider and Blikstein (2015) found that cross-recurrence is correlated with the learning gains. In a nutshell, gaze is coupled with cognition, and since gaze is coupled with dialogue, DUET methods constitute a powerful tool with which to quantitatively investigate the quality of collaboration. The observed correlations do not imply causality, but some studies show that displaying the gaze of one peer to the other, as a deictic gesture, increases team performance (Duchowski et al., 2004; Sharma, D’Angelo, Gergle, & Dillenbourg, 2016; Stein & Brennan, 2004; Van Gog, Jarodzka, Scheiter, Gerjets, & Paas, 2009; Van Gog, Kester, Niveltstein,

(2017). An application of extreme value theory to learning analytics: Predicting collaboration outcome from eye-tracking data. *Journal of Learning Analytics*, 4(3), 140–164. <http://dx.doi.org/10.18608/jla.2017.43.8>

Giesbers, & Paas, 2009; Van Gog & Scheiter, 2010). More importantly, these reported studies have mostly been conducted using ANOVAs, correlation tests, F- and t-tests and regressions, which assume that the data follow a normal distribution. We will show that the distribution tail of eye-tracking data (low frequency events) is quite different from the tail of normal distribution. Specifically, EVT hypothesizes that the events that occur in the tail of a distribution are more distinguishable than average behaviour. The next section, therefore, introduces the basics of EVT.

### 3 AN INTRODUCTION TO EXTREME VALUE THEORY

Extreme events are defined as those having low frequency and high severity (or impact). EVT is a branch of statistics that deals with modelling the occurrence and magnitude of such events. For instance, flood-walls are not built for average events but rather for rare and catastrophic occurrences. EVT for financial or insurance risk management looks at extreme events and concentrates on the risk of situations that might never have happened before (McNeil, Frey, & Embrechts, 2015). Such events (market crashes, insurance losses, etc.) are rare but very severe for companies, hence the need to model the deviations from the central tendencies in a different manner. Actually, the distribution of financial time series is known to be heavy-tailed. Therefore, EVT methods aim to model the tail with concepts described hereafter. For a comprehensive introduction, see Coles (2001), or see Chavez-Demoulin and Davison (2012) for a review of EVT for analyzing time series.

EVT is based on asymptotic results. Therefore, the data used to model events is a very small subset of the whole dataset (usually above the 90th or 95th quantile). The main advantages of using EVT<sup>1</sup> are as follows: First, it is based on the mathematical foundations that for any common distribution  $F$ , we can characterize the tail of  $F$  and can therefore understand the generating process of extreme events from any underlying distribution  $F$ .  $F$  can be any standard continuous distribution (normal, student, uniform, exponential, gamma, etc.); hence, EVT imposes no strong assumption upon the data generating processes, unlike ANOVAs. Second, when analyzing the dependence structure between two sequences of extreme events, the bivariate EVT context does not impose a linear shape of dependence as correlation requires (Sharma, Chavez-Demoulin, & Dillenbourg, 2016). Third, even if the theory is established for independent and identically distributed variables, it can be straightforwardly extended to the stationary context — the context we meet in eye-tracking and collaborative learning — or to the non-stationary context. Why is dual eye-tracking a stationary context? The gaze time-series are invariant of temporal-shifts, i.e., if we shift the time by a factor, the variability in the gaze patterns remain the same. Moreover, the gaze data at time  $t$  are not completely independent of where the person was looking at time  $t - 1$ , i.e., there exists an auto-correlation in the gaze data. Furthermore, we describe the advantages of EVT over general methods used in behavioural research:

- **Advantage of EVT over parametric models that assume normality of the data:** As previously mentioned, EVT does not assume any underlying distribution that generates the data. That is,

---

<sup>1</sup> Source: <http://www.bioss.ac.uk/people/adam/teaching/OR EVT/2007/node12.html>

(2017). An application of extreme value theory to learning analytics: Predicting collaboration outcome from eye-tracking data. *Journal of Learning Analytics*, 4(3), 140–164. <http://dx.doi.org/10.18608/jla.2017.43.8>

EVT can be applied to data from any standard continuous distribution (normal, student, uniform, exponential, gamma, etc.).

- Advantage of EVT over parametric models applied on the normalized data:** EVT offers a complementary viewpoint to look at the data, more particularly to look at the tail of the data distribution. This is justified because, often in the learning analytics context, the tail of the distribution is more informative than the body of distribution. This is illustrated by the real data of Figure 5. In that context, even if the normality of the transformed data hold, the parametric models applied on the data would not bring much information because there is no dependence structure to explore the average values (the points seem to be randomly spread in the middle quadrant of the plot containing the average values). More generally, when a group of students is interacting to accomplish a task, the upper tail of the joint distribution of temporal concentration (or lower tail of the joint distribution of their spatial entropy, like in Figure 5) actually represents the episodes during which the subjects are together focused in a high level of collaborative quality. The average joint values are less informative, probably containing other effects than collaboration. In such cases, the competitive performance of EVT approaches over parametric models, applied on the normalized data, emerges from the fact that EVT provides the correct tools to look at the extreme sequences of the data.
- Advantage of EVT over non-parametric models:** Both rely only on the assumption that the data are continuous. Many of the non-parametric methods used in learning analytics are hypothesis testing and provide one value (the p-value), which summarizes the data. Non-parametric forms can handle only low dimensional problems, which goes against the flow of big data. In general, in the (non-stationary) time series context, there is much more to gain from dynamic parametric models than from hypothesis testing. Because EVT is available for any common continuous distribution, it offers the advantages of parametric models like relying on likelihood, allowing formal inference, likelihood ratio-based hypothesis tests, and also takes into account non-stationary nature in the case of time series and covariate dependence. Note that non-parametric methods in the EVT context are also possible.

### 3.1 Univariate Case

Classical EVT considers two different approaches. The first approach provides the asymptotic behaviour of the maximum:

$$M_n = \max \{X_1, X_2, \dots, X_n\} \tag{1}$$

where  $X_1, X_2, \dots, X_n$  is an independent and identically distributed random sequence with distribution  $F$ . Suppose that we can find sequences of real numbers  $\{a_n > 0\}$  and  $\{b_n\} \{a_n > 0\}$  such that the sequence of normalized (or stabilized) maximum  $M_n^* = (M_n - b_n)/a_n$   $M_n^* = \frac{M_n - b_n}{a_n}$  converges in distribution.

A remarkable result states that the only possible distribution for the maximum is the generalized

extreme value (GEV) distribution:

$$G(z) = \exp \left[ - \left\{ 1 + \xi \left( \frac{z-\mu}{\sigma} \right) \right\}^{\frac{-1}{\xi}} \right] \tag{2}$$

where  $-\infty < \mu < \infty$  is the location parameter,  $\sigma > 0$  is the scale parameter, and  $-\infty < \xi < \infty$  is the shape parameter. This result is equivalent to the well-known central limit theorem (which provides a limiting distribution for the mean of any underlying distribution) but for the maximum. Concretely, in modelling extremes of a series of observed data  $x_1, x_2, \dots, x_n$ , we divide the data into  $m$  blocks of  $n$ . This gives us an observed series of block maxima  $m_{n,1}, m_{n,2}, \dots, m_{n,m}$  on which we fit a GEV, by maximum likelihood estimation, and get estimated location ( $\hat{\mu}$ ), shape ( $\hat{\sigma}$ ), and scale ( $\hat{\xi}$ ) parameters. The top panels in Figure 1 show an example of the selection of extreme events using the blockwise-maxima method for GEV model fitting. The second classical EVT approach (mathematically related to the first one) characterizes the tail of any continuous common distribution  $F$  and is referred to as the peaks-over-threshold (POT) approach. More precisely, it considers a model for the exceedances above some high threshold  $u$  that defines the tail of the distribution  $F$ . Under the POT approach it can be shown that:

- the number of exceedances above the threshold  $u$  arises according to a Poisson process with parameter  $\lambda$ , and independently,
- the exceedance size  $W = X - u$  follows a generalized Pareto distribution (GPD):

$$H(w) = 1 - \left( 1 + \frac{\xi w}{\bar{\sigma}} \right)^{\frac{-1}{\xi}} \tag{3}$$

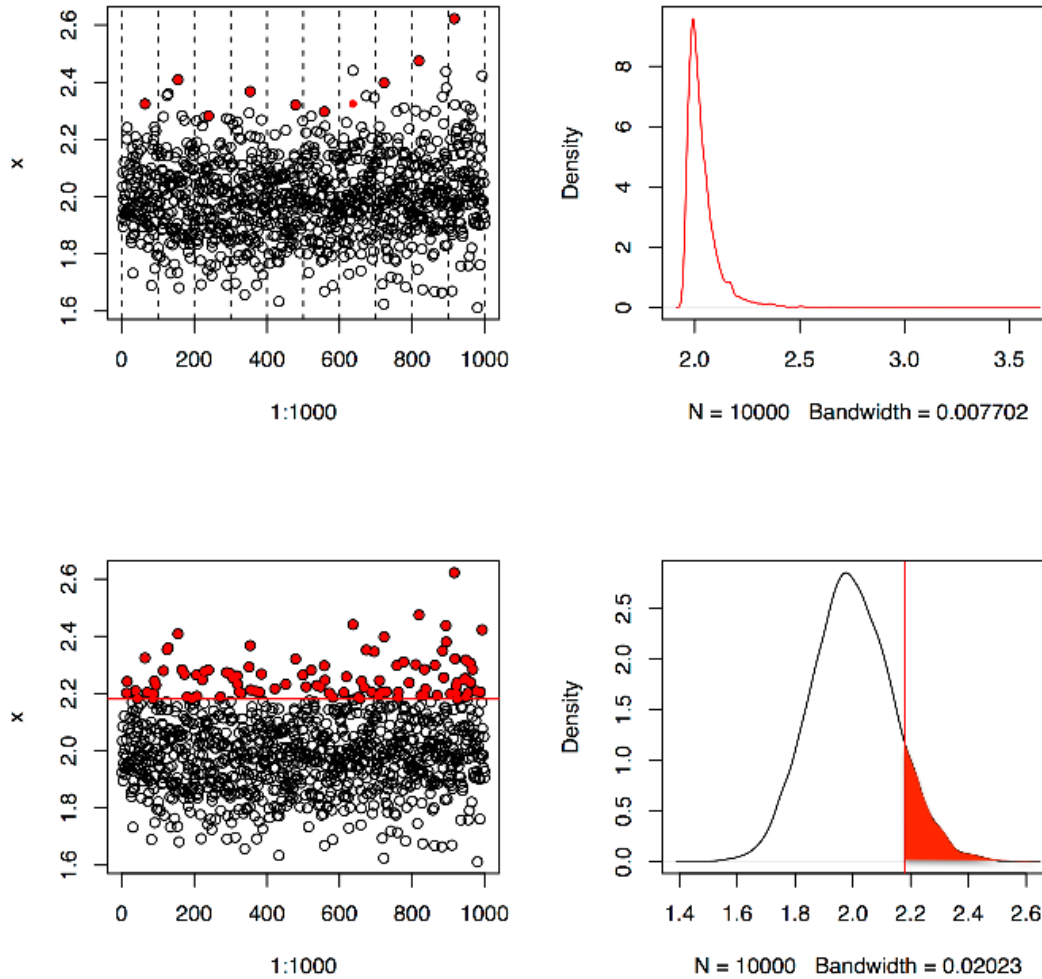
defined on  $\{w : w > 0 \text{ and } (1 + \xi w/\bar{\sigma}) > 0\}$ , where:

$$\bar{\sigma} = \sigma + \xi(u - \mu) \tag{4}$$

Essentially, parameters of the GPD (threshold excesses) can be determined by GEV (block maxima). The parameter  $\xi$ , which controls the shape of the tail of the distribution  $F$ , is the same for both GPD and GEV. In applications, the POT approach is more flexible than the block maxima approach and often allows more data (more than just one per block) and therefore leads to less uncertainty. As we can see in the top-left panel of Figure 1 (below), the number of points considered for modelling are the same as the number of blocks. On the other hand, the number of points in the bottom-left panel of Figure 1 is larger than when the POT method is used. Once we have determined the appropriate threshold, the parameter  $\lambda$  of the Poisson process and the GPD parameters  $\bar{\sigma}$  and  $\xi$  can be estimated by maximizing the likelihood function. The bottom panels in Figure 1 show an example of the selection of extreme events using the POT method for the GPD model fitting.



(2017). An application of extreme value theory to learning analytics: Predicting collaboration outcome from eye-tracking data. *Journal of Learning Analytics*, 4(3), 140–164. <http://dx.doi.org/10.18608/jla.2017.43.8>



**Figure 1: Top left: a random variable simulation and the blockwise-maxima. Top right: the density plot for one of the blocks; the red points show the maximum value of each block. Bottom left: the same random variable as in the top-left panel, the red horizontal line shows the threshold for the POT method, the red points are the points-over-threshold. Bottom right: the density plot for the whole distribution; the red vertical line shows the threshold for the POT method and denotes the beginning of the tail for the distribution; the red coloured area shows the tail, which corresponds to the red points in the bottom-left panel.**

The main practical use of such fitted models (GEV block maxima or POT) is the adequate calculation of the extreme quantile of  $F$ , that is, the quantile at a very high level. Using either the GEV or POT, we calculate a value, which has a very low probability of being exceeded in a given time period. This value is called the “return value,” a name inspired by environmental data in which the corresponding return period question is in how many months or years can it be expected that a value of the time series exceeds the same value again. The return value is set at a very high quantile, usually 95%, which means that there is only a 5% chance that a value will exceed the computed return value. In Section 4, we

(2017). An application of extreme value theory to learning analytics: Predicting collaboration outcome from eye-tracking data. *Journal of Learning Analytics*, 4(3), 140–164. <http://dx.doi.org/10.18608/jla.2017.43.8>

expose the calculation of the return level, and in Section 6 we see that the return level is actually effective for determining collaborative quality.

### 3.2 Bivariate Case

Another way of modelling collaboration with EVT is to use the gaze patterns from the two participants in a pair and analyze them as a bivariate time series. Given a bivariate random sample  $(X_1, Y_1), \dots, (X_n, Y_n)$ , EVT addresses the limiting behaviour of the component-wise maxima  $(M_{1,n}, M_{2,n})$ , that is, the respective maximum of the sequences  $\{X_i\}$  and  $\{Y_i\}$ ,  $i = 1, \dots, n$  as in (1).

The asymptotic theory of bivariate extremes deals with finding a non-degenerate bivariate distribution function (that can take more than two values)  $G$  such that, as  $n \rightarrow \infty$

$$Pr \left\{ (M_{1,n} - b_{1,n})/a_{1,n} \leq x, (M_{2,n} - b_{2,n})/a_{2,n} \leq y \rightarrow G(x, y) \right. \quad (5)$$

with sequences  $a_{l,n} > 0$  and  $b_{l,n} \in R$ ,  $l = 1, 2$ . If the limit (5) exists<sup>2</sup> and  $G$  is a non-degenerate distribution function, then  $G$  has the form:

$$G(z_1, z_2) = exp \left\{ - \left( \frac{1}{z_1} + \frac{1}{z_2} \right) A \left( \frac{z_2}{z_1 + z_2} \right) \right\} \quad (6)$$

The function  $A(\omega)$  defined as  $0 \leq \omega \leq 1$  is the so-called Pickands dependence function. The independence case corresponding to  $G(z_1, z_2) = exp\{-(1/z_1 + 1/z_2)\}$ , the Pickands function  $A(\omega)$ , measures the departure from independence. Complete dependence between the two series is reflected by  $A(1/2) = 0.5$ ; while at complete independence,  $A(1/2) = 1$ .

While analyzing the eye-tracking time series of two peers, the main practical use of the bivariate EVT is to measure extreme dependence, which is the probability of finding an extreme event in one time series, given that we observe an extreme event in the second time series. The two extreme events must occur at the same time, as the two dimensions in this bivariate space are the two gaze time series for the two peers. This probability is quantified as the tail-dependence between the two time series. The classical methods value typically used to measure the dependence between the two series is the correlation coefficient. The correlation coefficient is computed at the central tendencies, while the tail-dependence is, as in the case of return values, computed at a very high quantile. In Section 4, we use three different extremal dependence measures as complementary and interpretable ways for determining collaborative quality.

## 4 CONCEPTS

To apply EVT to our research question, predicting collaboration quality from DUET traces, we need to define a few variables.

<sup>2</sup> To simplify the representation and without loss of generality, we transform the data  $(X_i, Y_i)$  to  $(Z_{1i}, Z_{2i})$ ,  $i = 1, \dots, n$  with standard Fréchet margins so that  $Pr(Z_{il} \leq z) = exp\{-1/z\}$  for all  $z > 0$  and  $l = 1, 2$ .



(2017). An application of extreme value theory to learning analytics: Predicting collaboration outcome from eye-tracking data. *Journal of Learning Analytics*, 4(3), 140–164. <http://dx.doi.org/10.18608/jla.2017.43.8>

### 4.1 Gaze Visual Agitation (VA)

VA is defined as the coefficient of variance (CoV) of the fixation duration. Visual agitation for a given time window  $t$  is computed as follows:

$$VA_t = \frac{St. Dev. Fixation duration during time t}{Mean Fixation duration during time t} \tag{7}$$

In accordance with Richardson, Dale, and Tomlinson (2009), we chose a time window size of two seconds. The main reason for analyzing the variance of the fixation duration and not the fixation duration itself is the fact that the fixation duration is task-dependent. For instance, in a visual search task, the fixation durations will inherently be small, as the eyes would be constantly moving to search the target object, whereas in a task that requires deeper information processing, the fixation durations are higher. The task used in our experiment, drawing a concept-map task, lies in between: short fixation durations when peers search for a concept on the map versus longer fixations when they discuss the link between the two concepts. In order to keep various task episodes comparable, we use the scaled variance of the fixation duration. A low value of VA would mean relaxed gaze patterns while a high value could result from stress or fatigue.

### 4.2 Gaze Spatial Entropy (SE)

SE measures the spatial distribution of the gaze of each peer. To compute SE, we first define a 100-pixel-by-100-pixel grid over the screen and we compute for each peer the proportion of gaze fixations located in each grid cell (Figure 2). This results in a proportionality matrix and the SE is computed as the Shannon entropy of this 2-dimensional vector. The spatial entropy is also task-independent, as it can be computed for any task, but the interpretation of the entropy values might be dependent on the visual stimuli. A low value of SE would mean that the subject is concentrating on a few elements on the screen, while a high SE value would depict a wider focus size.

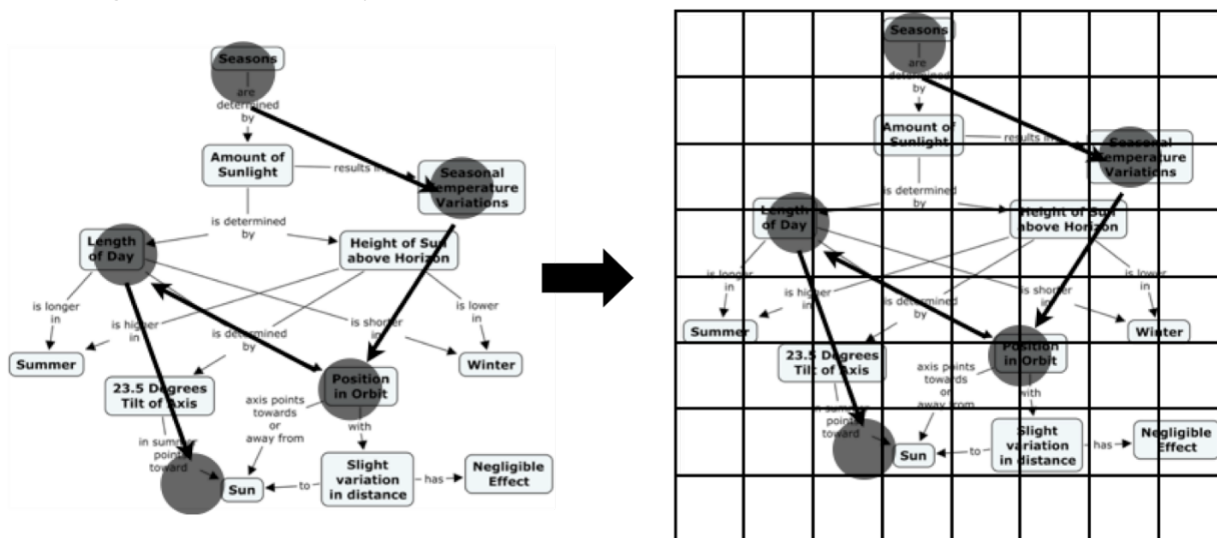


Figure 2: The process of computing entropy. The image on the left shows the exemplar concept-map and gaze patterns (grey circles and arrows). The image on the right shows the placement of the grid.

### 4.3 Return Levels: Univariate Extremes

The return level is the quantile at a high level (above 90% for example) of the data distribution. Why do we not simply calculate this quantile from the distribution of our entire dataset? We could do this, but small discrepancies in the estimation of the body distribution would lead to large errors in the estimation of the quantiles in the tail. The POT model presented in Section 3 is the mathematically correct way to estimate such high quantiles and in practice leads to more accurate estimation. The EVT estimation also brings information about how heavy is the tail of the distribution  $F$ ; that is, how large are the extremes that distribution  $F$  can generate? This information is provided by the value of the shape parameter  $\xi$  in (2) or (3): as  $\xi$  becomes larger, the tail of  $F$  becomes heavier. We do not explore this feature further in this paper because as with any other modelling approach, just from the set of estimated parameters of location  $\mu$ , scale  $\sigma$  or  $\bar{\sigma}$ , and shape  $\xi$ , it is cumbersome to explain and compare the different models. Hence, we use the return level, calculated using the model parameters, which has a valuable interpretation.

As mentioned in Section 3, the return value (say, calculated at the 95% quantile), symbolizes the measure of the (unseen) extreme event with a 5% probability that the actual (unseen) event exceeds this value. In what follows, we derive the return level calculation from the POT model above a threshold  $u$ . We recall that the underlying variable is denoted  $X$  and that the exceedances occurrence arrives according to a Poisson process with parameter  $\lambda$ , and the exceedance size  $W = X - u$  follows a GPD denoted as  $H$  in (3) with parameters  $(\bar{\sigma}, \xi)$ . For  $x > u$ , we have:

$$\begin{aligned} Pr(X > u | x > u) &= Pr(X - u > x - u | X > u) \\ &= 1 - Pr(W \leq x - u | X > u) \\ &= 1 - H(x - u) \\ &= \left\{ 1 + \xi \left( \frac{x-u}{\bar{\sigma}} \right) \right\}^{-\frac{1}{\xi}} \end{aligned}$$

It follows that

$$Pr(X > x) = Pr(x > u) \left\{ 1 + \xi \left( \frac{x-u}{\bar{\sigma}} \right) \right\}^{-\frac{1}{\xi}} \tag{8}$$

Hence, the return level  $x_p$  or extreme quantile at the percentile  $p$  (large) is the solution of

$$Pr(x > u) \left\{ 1 + \xi \left( \frac{x-u}{\bar{\sigma}} \right) \right\}^{-\frac{1}{\xi}} = 1 - p \tag{9}$$

so that,

$$x_p = \left\{ \begin{array}{ll} u + \frac{\bar{\sigma}}{\xi} \left[ \left\{ \frac{Pr(X>u)}{1-p} \right\}^{\xi} - 1 \right] & \text{for } \xi \neq 0 \\ u + \bar{\sigma} \log \left\{ \frac{Pr(X>u)}{1-p} \right\} & \text{for } \xi = 0 \end{array} \right\} \tag{10}$$

(2017). An application of extreme value theory to learning analytics: Predicting collaboration outcome from eye-tracking data. *Journal of Learning Analytics*, 4(3), 140–164. <http://dx.doi.org/10.18608/jla.2017.43.8>

In a non-mathematical way, the return level  $x_p$  is the value at which the probability of exceeding this value is equal to  $1 - p$ . We obtain the estimated return level (10) by fitting the POT model to the exceedance data, estimating the probability of exceeding the threshold,  $Pr(X > u)$ , using the Poisson model and replacing the parameters  $\sigma$  and  $\xi$  with their maximum likelihood estimates.

Is EVT overkill, or is it really necessary to analyze the two variables that we have defined, visual agitation and spatial entropy? Figure 3 uses Q–Q plots for comparing the distribution of these two variables with a normal distribution. Both plots show a heavy tail for low frequency values of spatial entropy (left plot) and visual agitation (right plot), respectively. This justifies the use of sophisticated EVT methods to process these tails. We will therefore compare the return levels calculated for the two participants. Similar return levels would depict a higher amount of temporal concordance. In Section 6, we will see that comparing return levels indeed provides an accurate (and interpretable) way of discriminating high and low collaboration quality.

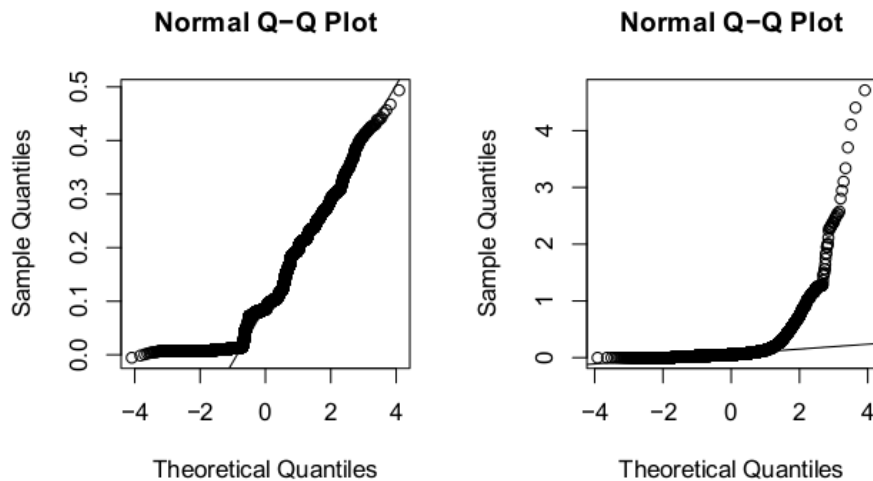


Figure 3: Q–Q plots of Spatial Entropy (left) and Visual Agitation (right) defined in Section 4.

#### 4.4 Three Measures of Extremal Dependence: Bivariate Extremes

Estimating dependence between the two partners in a pair’s extremal behaviour provides some complementary information about the peers’ concordance. We first introduce the extremal coefficient

$$\vartheta = 2A(1/2) \tag{11}$$

where  $A$  is the Pickands function mentioned in Section 3. Thus,  $\theta \in [1, 2]$ , and it can be conveniently interpreted as the effective number of independent series; the case  $\theta = 2$  means that the two series are independent and we therefore get complete independence. The case  $\theta = 1$  means that the effective number of independent series is 1, and therefore we get complete dependence.

The two other extremal dependence measures we consider come from conventional multivariate extreme value theory, characterizing two classes of extreme value dependence: asymptotic dependence

(2017). An application of extreme value theory to learning analytics: Predicting collaboration outcome from eye-tracking data. *Journal of Learning Analytics*, 4(3), 140–164. <http://dx.doi.org/10.18608/jla.2017.43.8>

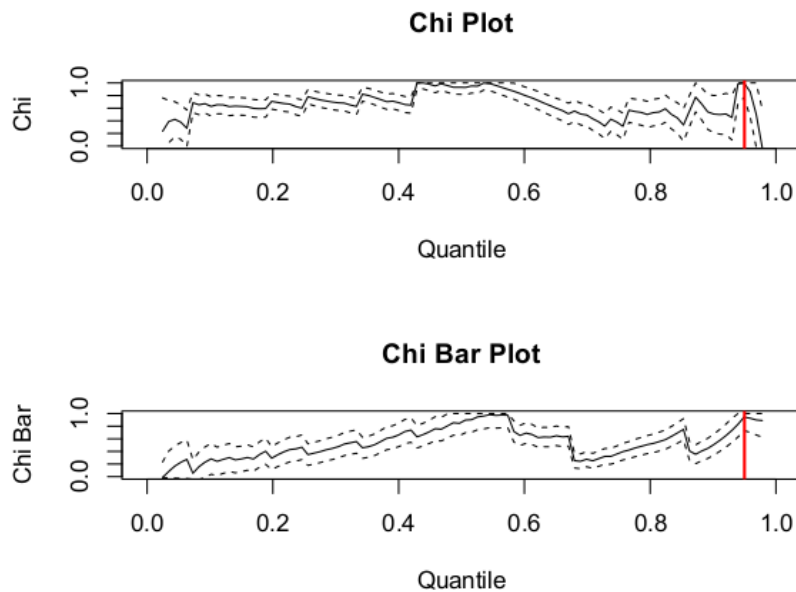
and asymptotic independence, which characterizes the behaviour of variables as they become more extreme. In this context, we consider the coefficient of extremal dependence

$$\chi = \lim_{z \rightarrow \infty} Pr(Z_1 > z | Z_2 > z) \tag{12}$$

The limit value  $\chi \in [0, 1]$  is strictly positive when a large value of  $Z_2$  leads to a non-zero probability of as large as value  $Z_1$ . In other words,  $\chi$  is the tendency for one variable to be large given that the other is large. This means that the only possibility for asymptotic independence is when  $\chi = 0$ . When  $\chi > 0$ , the variables are asymptotically dependent. In that context, we define, as a second extremal coefficient, the conditional probability

$$\bar{\chi} = \lim_{z \rightarrow 0} Pr(Z_1 \leq z | Z_2 \leq z) \tag{13}$$

From this we see that  $\bar{\chi} = 1$  means perfect dependence between the two series while  $\bar{\chi} = 0$  implies independence. The coefficient  $\bar{\chi}$  is therefore a measure of dependence for the class of asymptotically independent models. In our context,  $\chi$  tells us the level of asymptotic dependence, and  $\bar{\chi}$  tells us about the strength of the asymptotic dependence. In practice, as (12) and (13) are limits, we set a value of  $z$  typically at a very high quantile for (12) and very low one for (13), referred to as  $z \times 100$  percentile for (12) and taking the  $(1 - z) \times 100$  percentile for (13), as shown in the results in Section 6.



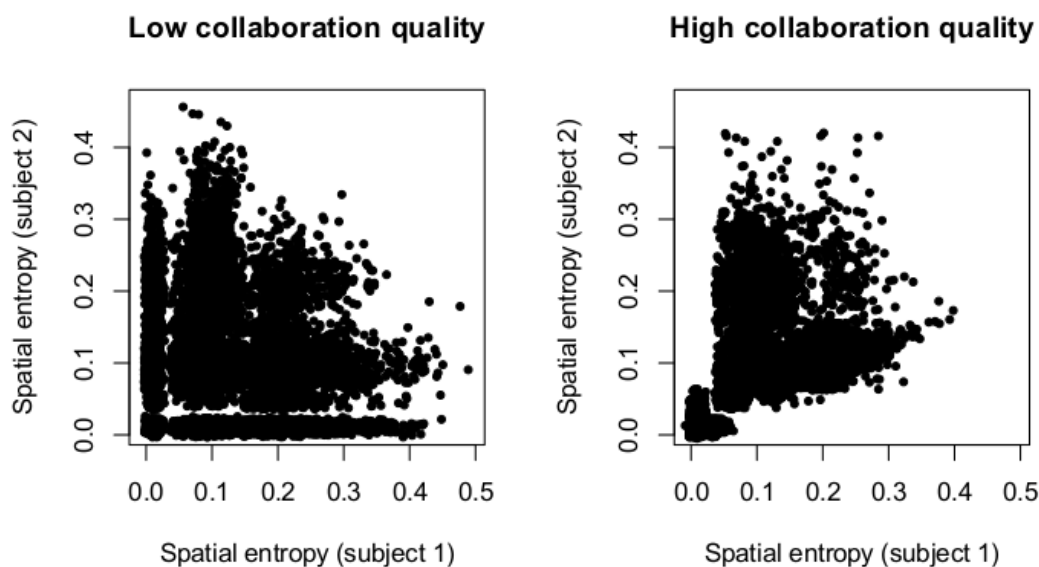
**Figure 4: Example illustrating the determination of the coefficient of extremal dependence  $\chi$  and the strength of dependence  $\bar{\chi}$  for the visual agitation of a pair. The dashed lines represent the 95% confidence intervals for  $\chi$  and  $\bar{\chi}$ . The tail-dependence and its strength is determined by the values at the higher quantiles (typically between 95% and 99%). The red lines correspond to 95%.**

Figure 4 shows an example illustrating the determination of the coefficient of extremal dependence  $\chi$  and the strength of dependence  $\bar{\chi}$  for the spatial entropy of a pair. Why do we calculate  $\chi$  and  $\bar{\chi}$  for all

(2017). An application of extreme value theory to learning analytics: Predicting collaboration outcome from eye-tracking data. *Journal of Learning Analytics*, 4(3), 140–164. <http://dx.doi.org/10.18608/jla.2017.43.8>

the quantiles? This is just an empirical method, and we are only interested in the highest quantile values.

Again, is bivariate EVT overkill, or is it really necessary to analyze the variables that we have defined, visual agitation and spatial entropy? Figure 5 shows that the dependence structure between the spatial entropy of the two peers is far from linear (for both low and high collaborative quality pairs). In such a case, a Pearson correlation would lead to erroneous conclusions. This leads to the development of more sophisticated methods to adequately model dependence structure; see, for instance, Sharma et al. (2017).



**Figure 5: Scatterplots of spatial entropy between the peers with low (left panel) and high (right panel) quality of collaboration.**

## 5 EXPERIMENT

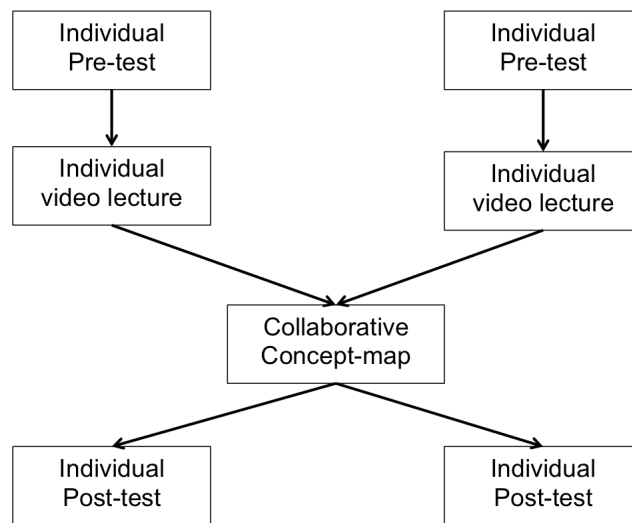
The EVT framework presented above provides a new method for analyzing the dual eye-tracking data. The research question we specifically address is the following: **Do extreme values from gaze episodes predict the quality of collaboratively produced concept maps better than central trends?**

To answer this question, we conducted an experiment with 66 master’s students from École Polytechnique Fédérale de Lausanne who participated in the present study. There were 20 females among the participants. The participants were each compensated with 30 Swiss francs for their participation in the study. The flow of the experiment is shown in Figure 6.

Upon their arrival in the laboratory, the participants signed a consent form. Then they took an individual pre-test on the basics of neuronal transmission. Then the participants individually watched two videos about “resting membrane potential.” Next, they created a collaborative concept-map using IHMC CMap

(2017). An application of extreme value theory to learning analytics: Predicting collaboration outcome from eye-tracking data. *Journal of Learning Analytics*, 4(3), 140–164. <http://dx.doi.org/10.18608/jla.2017.43.8>

tools.<sup>3</sup> Finally, they took an individual post-test. The two videos were taken from “Khan Academy.”<sup>4,5</sup> The total length of the videos was 17 minutes. It is worth mentioning that the teacher was not physically present during the videos. The participants came to the laboratory in pairs. While watching the videos, the participants had full control over the video player without any time constraint. The collaborative concept-map phase was 10–12 minutes long. During that time participants could talk to each other while their screens were synchronized, i.e., peers were able to see each other’s actions. Both the pre-test and the post-test contained true–false questions.



**Figure 6: Schematic representation of the different phases of the experiment.**

### 5.1 Quality of Collaboration

The final concept-map was compared with the concept-map created by the two experts. The pair received a score using the following rules: 1) one mark for each correct connection between two concepts, 2) one mark for each correct label of the edge between two concepts, 3) half a mark for each partially correct label of the edge between two concepts. The pairs were then divided into two levels based on the concept-map score using a median split. Why do we consider this as a measure of collaboration quality? The reason rests in the work of Jermann, Mullins, Nüssli, and Dillenbourg (2011), Jermann and Nüssli (2012), and Kahrimanis, Chounta, and Avouris (2010), who showed that the actions/task-based outcome is often correlated with the collaboration quality. Hence, our assumption about having the collaborative product quality as a proxy of collaboration quality is grounded in previous findings. As Wise and Shaffer (2015) suggest, “...theory plays an ever-more critical role in analysis,” so using these supports from the literature, we can proceed with the aforementioned assumption.

<sup>3</sup> [CMap tools](#)

<sup>4</sup> [Resting Membrane Potential-Part 1](#)

<sup>5</sup> [Resting Membrane Potential-Part 2](#)

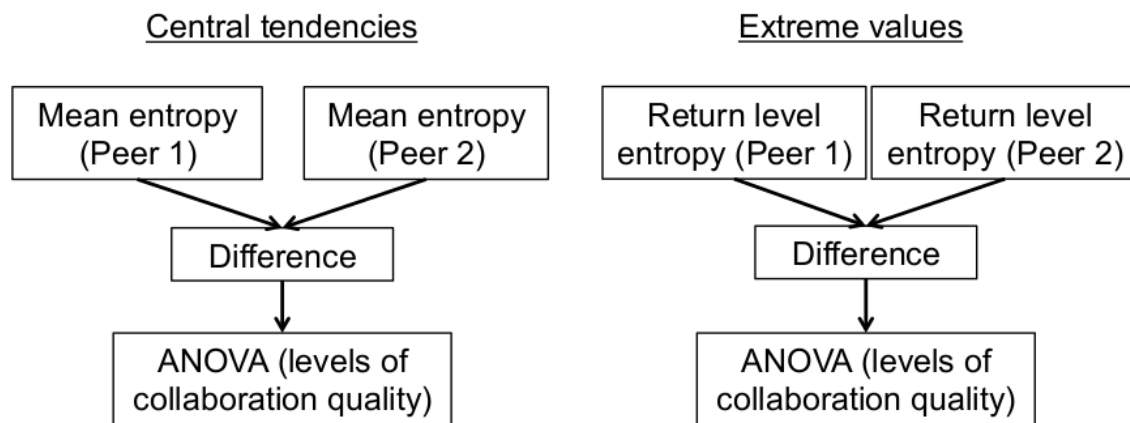


## 6 RESULTS

### 6.1 Univariate Extremes

Recall the question we address in this paper: **Does EVT reveal differences that central trends failed to reveal?**

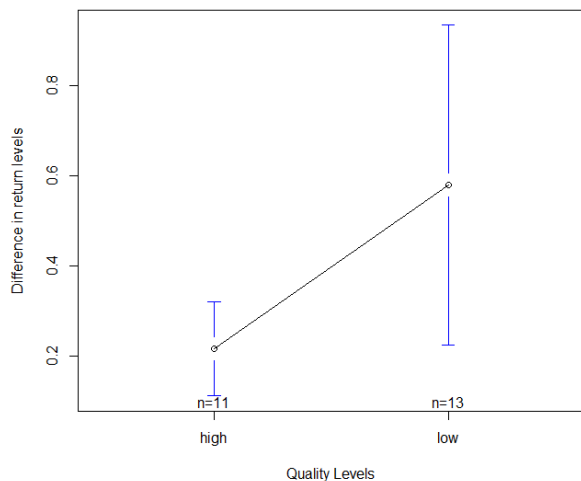
Figure 7 shows the pipeline for data processing. Let us begin with the central trends approach. If we compare the difference in the average levels of entropy of the peers, we observe no significant differences between high- and low-quality pairs. An ANOVA shows no significant difference in the average entropy difference for the peers with high and low collaboration quality ( $F[1,21.48] = 0.01$ ,  $p$ -value = .93, Figure 8d). The same lack of difference is found with the visual agitation ( $F[1,22] = 1.73$ ,  $p$ -value = .20, Figure 8c).



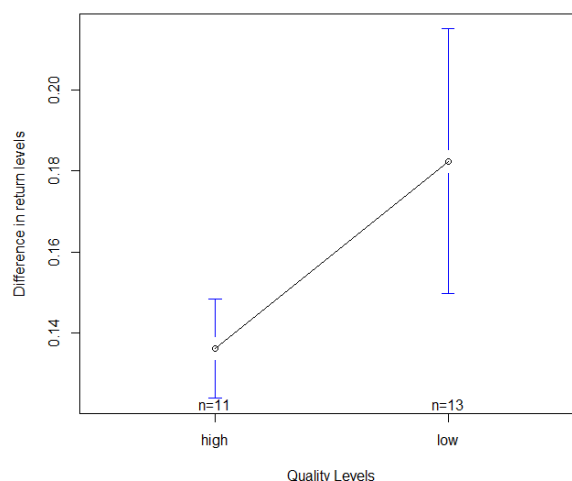
**Figure 7: The pipeline for univariate data-processing.**

Now, we compare the previous results with those provided by EVT. We estimated the return level (10) at percentile  $p$ . To keep enough data, we set  $p = 90$  for visual agitation and  $p = 95$  for spatial entropy. The reason for setting  $p = 90$  for visual agitation is to have enough data points to fit a GEV or POT. The difference between peers in terms of return levels tells us about their synchronicity. The difference between peers in return levels for visual agitation is lower for high-quality pairs than for low-quality pairs ( $F[1,14.08] = 4.92$ ,  $p$ -value = .04, one-way ANOVA without assuming equal variances). Similarly, the difference between peers in return levels for spatial entropy is also lower for high-quality pairs ( $F[1,15.15] = 8.39$ ,  $p$ -value = .01, one-way ANOVA without assuming equal variances). Figures 8a and 8b show the means and confidence intervals for the difference in the return levels for visual agitation and spatial entropy respectively. In other words, both for agitation and entropy, the extremes occur with higher synchronicity for the high-quality pairs than for the low-quality pairs.

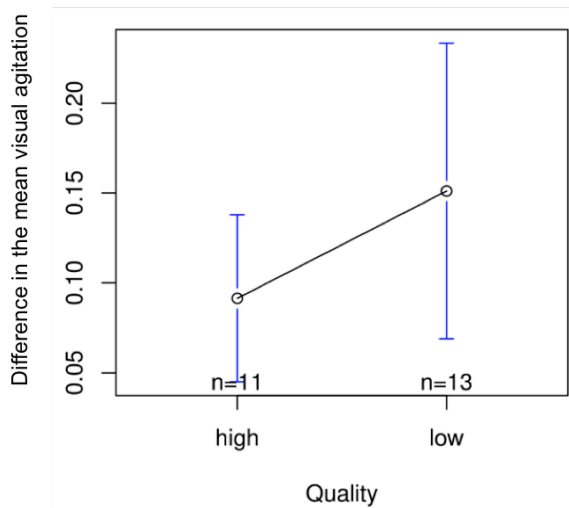
(2017). An application of extreme value theory to learning analytics: Predicting collaboration outcome from eye-tracking data. *Journal of Learning Analytics*, 4(3), 140–164. <http://dx.doi.org/10.18608/jla.2017.43.8>



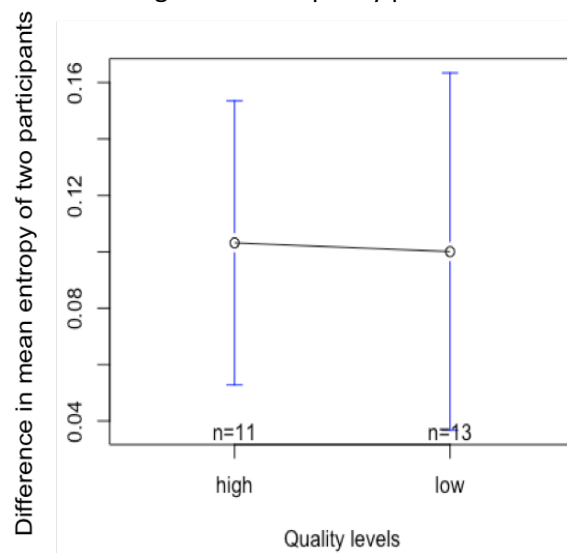
(a) Means and confidence intervals (blue bars) for the difference in the estimated return levels (10) at 90 percentile for visual agitation, for high- and low-quality pairs.



(b) Means and confidence intervals (blue bars) for the difference in the estimated return levels (10) at 90 percentile for spatial entropy, for high- and low-quality pairs.



(c) Means and confidence intervals (blue bars) for the difference in the mean values for visual agitation, for high- and low-quality pairs.



(d) Means and confidence intervals (blue bars) for the difference in the mean values for spatial entropy, for high- and low-quality pairs.

**Figure 8: Results: Univariate extremes**

## 6.2 Bivariate Extremes

We again compare the two methods: Does EVT reveal differences (of dependencies among peers) that central trends did not?

Let us start with standard correlations. If we compute the correlation between the spatial entropy of two peers, we can see in both Figures 10c and 10d, that we cannot learn anything from the average values (the body of the distribution), and the Pearson correlation/linear model does not make sense here. This might lead to false interpretations of the underlying collaborative processes.

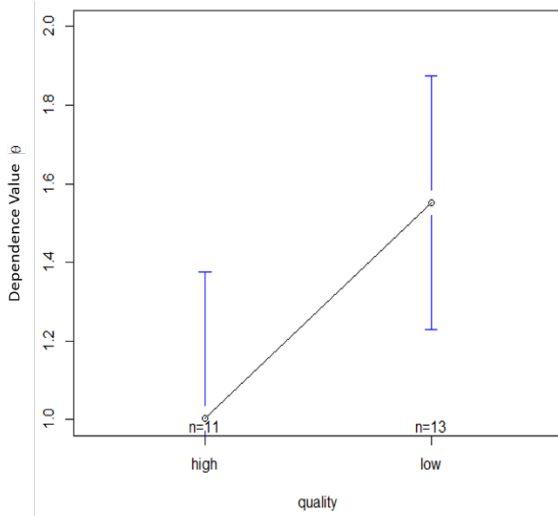
Let us now compare the EVT approach to the bivariate time series. To estimate their extremal dependence, we start by estimating the extremal coefficient  $\theta$  as in (11) between the variables for the two peers. We observe that high-quality pairs have a higher dependence for visual agitation than low-quality pairs ( $F[1,22] = 6.07$ ,  $p\text{-value} = 0.02$ , Figure 9a). Similarly, high-quality pairs have a higher level of dependence in visual entropy than low-quality pairs, with the difference being even more significant ( $F[1,22] = 7.65$ ,  $p\text{-value} = 0.01$ , Figure 9b). The scales on the y-axes for Figures 9a and 9b are inverted. As we mentioned in Section 4.4, complete dependence is reflected by  $\theta = 1$ , whereas complete independence is reflected by  $\theta = 2$ .

Next, we estimate the level  $\chi$  defined in (12) and strength  $\bar{\chi}$  defined in (13) of the extremal dependence. We observe a higher extremal dependence (calculated at the 95% quantile) between the visual agitation of peers for pairs with high collaboration quality ( $F[1,22] = 9.19$ ,  $p\text{-value} = 0.006$ , Figure 11a). Moreover, we observe an even more significant difference in the strength of the extremal dependence (calculated at the 95% quantile) in favour of the pairs with high collaboration quality ( $F[1,22] = 11.71$ ,  $p\text{-value} = 0.002$ , Figure 11c).

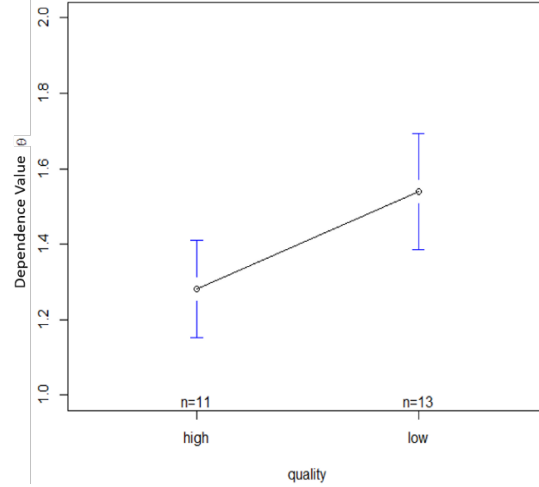
Regarding spatial entropy, we observe effects similar to visual agitation. There is a higher extremal dependence (calculated at the 95% quantile) between the spatial entropy of peers with high collaboration quality ( $F[1,22] = 6.31$ ,  $p\text{-value} = 0.01$ , Figure 11b). Similar to the case of visual agitation, we observe an even more significant difference in the strength of extremal dependence (calculated at the 95% quantile) for the pairs with high collaboration quality ( $F[1,22] = 14.28$ ,  $p\text{-value} = 0.001$ , Figure 11d).

There is a higher ( $\chi$ ) and stronger ( $\bar{\chi}$ ) (calculated at the 95% quantile) extremal dependence for both visual agitation (Figure 10a) and spatial entropy (Figure 10b) for the high-quality pairs than the low-quality pairs. We observe a clear separation, in the 2-dimensional space of  $\chi$  and  $\bar{\chi}$ , between the high- and low-quality pairs (with three and one exception for visual agitation and spatial entropy, respectively). As we observe in the case of temporal univariate return levels, the difference is more evident in the case of spatial entropy than in the case of visual agitation.

(2017). An application of extreme value theory to learning analytics: Predicting collaboration outcome from eye-tracking data. *Journal of Learning Analytics*, 4(3), 140–164. <http://dx.doi.org/10.18608/jla.2017.43.8>

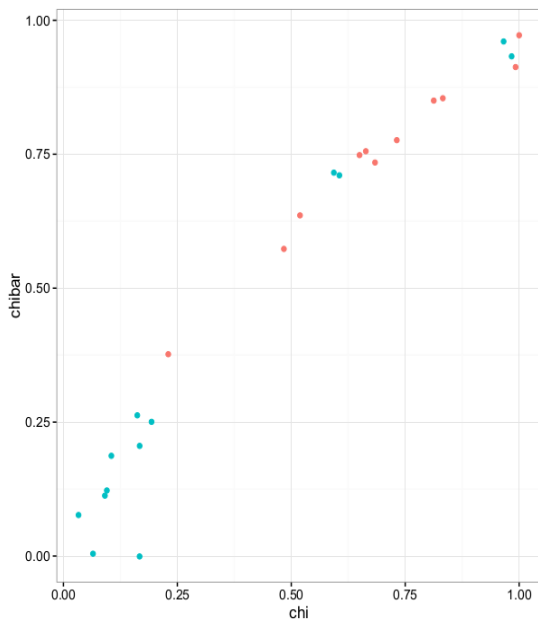


(a) Means and confidence intervals (blue bars) for the estimated extremal coefficient  $\theta$  for VA of the participants, for high- and low-quality pairs.

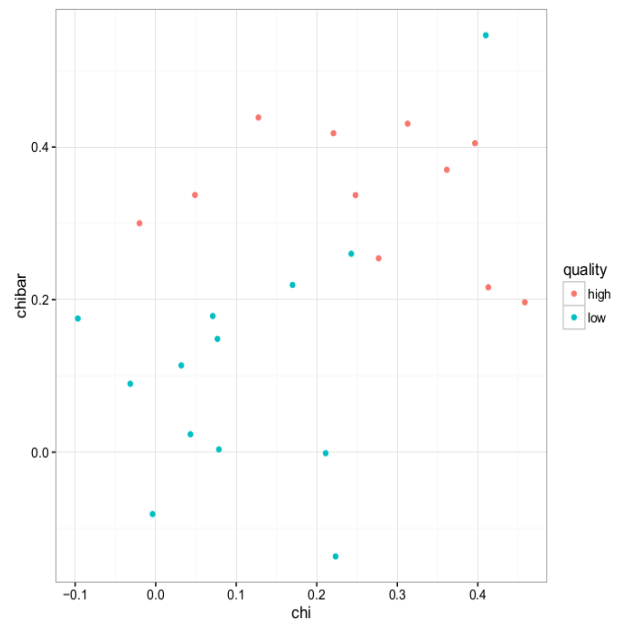


(b) Means and confidence intervals (blue bars) for the estimated extremal coefficient  $\theta$  for SE of the participants, for high- and low-quality pairs.

**Figure 9: Bivariate extremes: Dependence measures.**

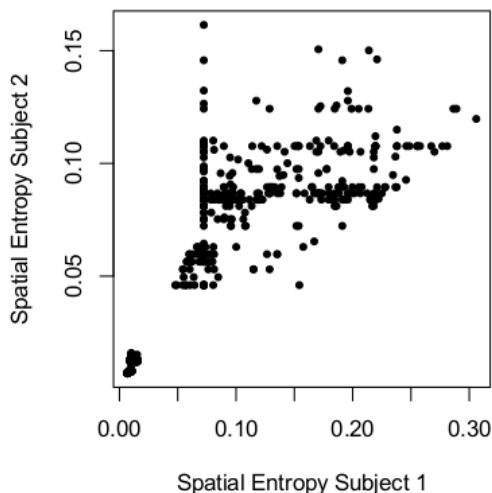


(a) Coefficient  $\chi$  and strength  $\bar{\chi}$  of extremal dependence for VA for high (red points) and low (blue points) collaboration quality pairs.

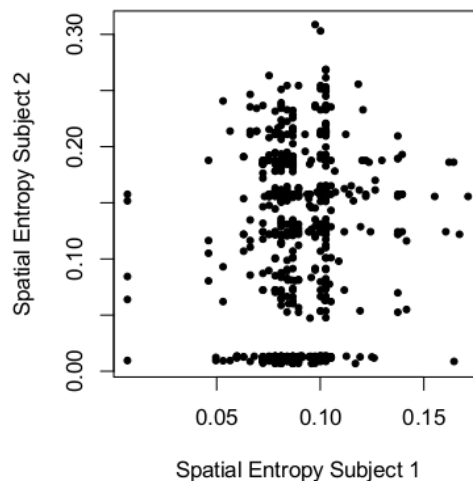


(b) Coefficient  $\chi$  and strength  $\bar{\chi}$  of extremal dependence for SE for high (red points) and low (blue points) collaboration quality pairs.

(2017). An application of extreme value theory to learning analytics: Predicting collaboration outcome from eye-tracking data. *Journal of Learning Analytics*, 4(3), 140–164. <http://dx.doi.org/10.18608/jla.2017.43.8>

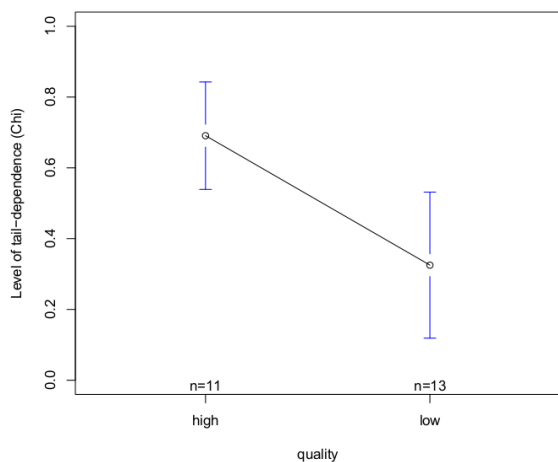


(c) SE values for peers in a high-quality pair. The correlation does not reflect the true relationship, as there is no linear relation between the SE values for peers.

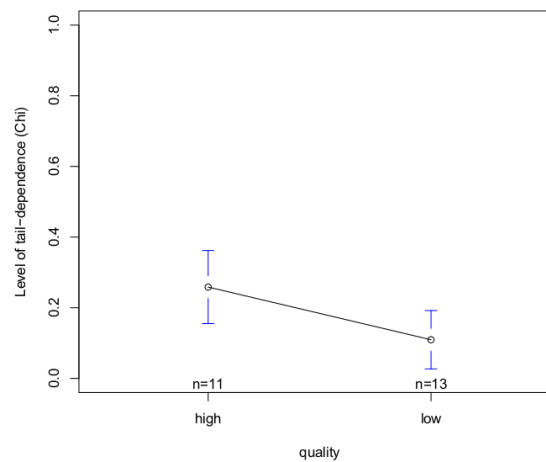


(d) SE values for peers in a low-quality pair. The correlation does not reflect the true relationship, as there is no linear relation between the SE values for peers.

**Figure 10: Results: Bivariate extremes, extremal coefficient, and tail dependence.**



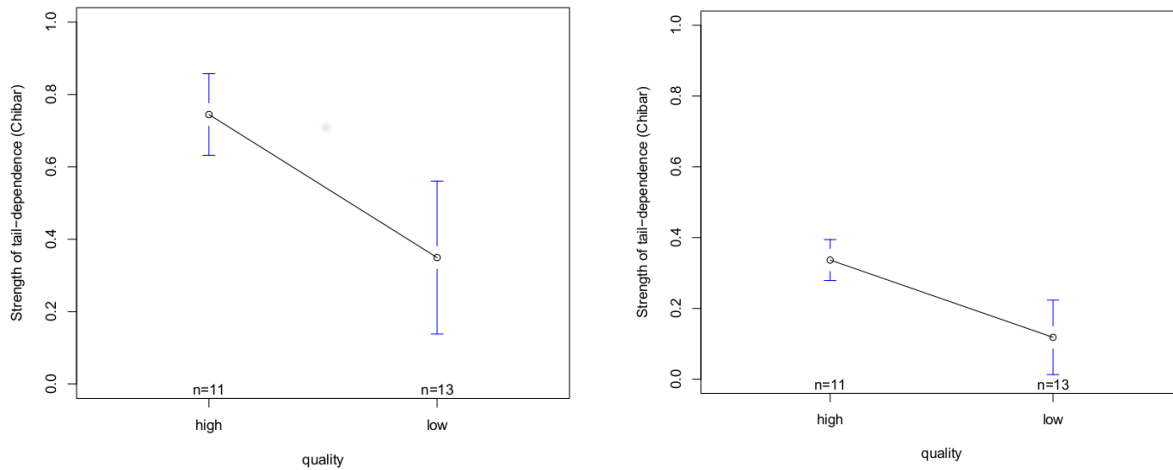
(a) Means and confidence intervals (blue bars) for the estimated level of extremal dependence  $\chi$  in the visual agitation of the participants, for high- and low-quality pairs.



(b) Means and confidence intervals (blue bars) for the estimated level of extremal dependence  $\chi$  for spatial entropy of the participants, for high- and low-quality pairs.

**Figure 11. Results: Bivariate extremes, levels, and strength of tail dependence**

(2017). An application of extreme value theory to learning analytics: Predicting collaboration outcome from eye-tracking data. *Journal of Learning Analytics*, 4(3), 140–164. <http://dx.doi.org/10.18608/jla.2017.43.8>



(c) Means and confidence intervals (blue bars) for the estimated strength of extremal dependence  $\bar{\chi}$  in the visual agitation of the participants, for high- and low-quality pairs.

(d) Means and confidence intervals (blue bars) for the estimated strength of extremal dependence  $\bar{\chi}$  in the visual agitation of the participants, for high- and low-quality pairs.

**Figure 11. Results: Bivariate extremes, levels, and strength of tail dependence.**

## 7 DISCUSSION

Does EVT provide interesting findings compared to statistical methods based on central trends?

Let us first address this question in the univariate context. The comparison of mean values of visual agitation or spatial entropy did not reveal any difference between high-quality and low-quality pairs. On the contrary, EVT revealed that high-quality pairs have a significantly smaller difference of return levels for both variables. This shows that during extreme episodes of collaboration there exists a higher amount of “togetherness” among the participants in high-quality pairs.

The bivariate context is even more interesting. The three tail dependence coefficients we used measure dependence between the extremes of visual agitation and spatial entropy in a time series. More specifically, from the extremal coefficient  $\theta$  we learn the effective number of independent series: for high-quality pairs,  $\theta \approx 1$ , meaning that the time series of one peer, for both variables, suffices to explain (or describe) the extremes of the other peer. This highlights an extreme “togetherness” in collaboration between the two participants of the pair.

The dependence measures  $\chi$  and  $\bar{\chi}$  play a role similar to the Pearson correlation, but they avoid the drawbacks of standard correlation (not robust to outliers, restricted to linear dependence structure, spoiled by other effects affecting the body of the distribution). The extremal dependence measures  $\chi$  and  $\bar{\chi}$  focus on the extreme values of the two variables. Similarly to the interpretation of correlation, large values of  $\hat{\chi}$  and  $\hat{\bar{\chi}}$  indicate a strong dependence between their episodes of high VA and SE. The fact



(2017). An application of extreme value theory to learning analytics: Predicting collaboration outcome from eye-tracking data. *Journal of Learning Analytics*, 4(3), 140–164. <http://dx.doi.org/10.18608/jla.2017.43.8>

that the bivariate tail-dependence is higher and stronger for the high-quality pairs confirms the univariate findings.

Using the bivariate space formed by the same gaze measure for both participants in the pair (both for VA and SE), we eliminate the need for grouping (averaging or grouping the individual measures in a regression model) the peer measures into pair variables.

## 7.1 Why Does EVT Work?

One reason EVT works is that, unlike standard methods that suffer from the difference between the assumed underlying distribution and the actual distribution, EVT properly models the tail (of any common distribution) using the correct model (POT or GEV block maxima). Second, when we use the extreme episodes, we focus only on the moments that might reflect the episodes during which the collaborators are most likely to be “together.” Then, by focusing on extreme collaboration episodes, we remove the noise that could have prevented classical methods from differentiating the collaboration quality levels. This fact is also evident in Figures 10c and 10d. Correlation does not reflect the correct relation between the SE for the two participants.

However, why could we not take the top 5% quantile and perform an ANOVA on those values? A very simple answer is that the main assumption for ANOVA is that the values should follow a normal distribution, and it is mathematically proven that the tail of any distribution, which is normal in the case of ANOVA, does not follow the distribution. Instead, it follows the GPD. Hence, it would be statistically wrong to perform an ANOVA on such variables. Could we simply normalize the data and then perform the ANOVA? This could lead to a problem as we completely ignore many other properties of data (e.g., skew and kurtosis) while normalizing the data. Thus, key aspects of the data generation process might be hidden or removed. EVT provides a method that assumes no underlying distribution regarding the data generating process, unlike other classical methods. This removes the need to force the data to follow any given statistical distribution.

## 7.2 When to use EVT?

EVT offers the correct way (in the sense that it is based on mathematical foundations) to analyze abnormal data (in the sense of data far from the average values). The EVT theory for the largest values or peaks-over-threshold or bivariate case exposed in the paper is available for any underlying continuous distribution. It should be used when analyzing the tail distribution (for any kind of continuous distribution) as a complementary exploration of the data, or when traditional methods fail or are uninformative, either because the assumptions required by these methods (like the linear model/Pearson correlation) based on linear dependence between the two variables are violated or nearly violated or because the average values on which all these (parametric or non-parametric) methods are based do not contain the relevant information of interest, being therefore less predictive. For example, when a student is writing in a graphical table, the extreme values of her time series of writing speed/pressure are her abnormal sequences (in the sense of departure from her standard

(2017). An application of extreme value theory to learning analytics: Predicting collaboration outcome from eye-tracking data. *Journal of Learning Analytics*, 4(3), 140–164. <http://dx.doi.org/10.18608/jla.2017.43.8>

measures) and relate to her episodes of stress. Another example, when a teacher looks at the exams to infer the heterogeneity of the class, she cannot just be satisfied by a robust measure of the variability of marks. She has to carefully consider the worst and the best marks (the extremes) as the limits of the class heterogeneity frame. Neglecting the worst and the best would not only mean neglecting some students (who probably have an important impact on the class) but also neglecting relevant information. Furthermore, while analyzing trace data (for example, click-streams), although the theory is not established for the discrete case, it is typically used to count variables, like Poisson variables, because of their approximation by continuous distribution.

## 8 CONCLUSION

It is easy to understand that a statistical model that predicts a rise in water level of 5 metres has more social relevance than a model that predicts a rise of 5 centimetres. In education, this approach is less intuitive. Typically, a teacher would care for the average level of his class and try to cope with its heterogeneity. It is hence very counter-intuitive that EVT reaches a higher discriminative power than methods based on central trends. In sciences, what is counter-intuitive is always interesting. However, we should not forget that the extreme values are not outliers but extreme time episodes during collaboration, which is less counter-intuitive. If a teacher monitors a classroom with several teams, (s)he would probably be also attracted by “extreme” episodes; for instance, when peers do not speak at all or when they shout at each other. In our experiment, the raw data is not dialogue but gaze patterns, and at this point nothing proves that similar results would be obtained with other behavioural traces. We do not claim that EVT should replace other statistical methods used in learning analytics, but rather that it expands the range of tools available to learning scientists. By using it across multiple learning contexts, we will learn when and why it brings more discriminative power than methods based on central trends.

## REFERENCES

- Abernethy, B., & Russell, D. G. (1987). The relationship between expertise and visual search strategy in a racquet sport. *Human Movement Science*, 6(4), 283–319. [http://dx.doi.org/10.1016/0167-9457\(87\)90001-7](http://dx.doi.org/10.1016/0167-9457(87)90001-7)
- Charness, N., Reingold, E. M., Pomplun, M., & Stampe, D. M. (2001). The perceptual aspect of skilled performance in chess: Evidence from eye movements. *Memory & Cognition*, 29(8), 1146–1152. <http://dx.doi.org/10.3758/BF03206384>
- Chavez-Demoulin, V., & Davison, A. C. (2012). Modelling time series extremes. *REVSTAT: Statistical Journal*, 10(1), 109–133.
- Cherubini, M., & Dillenbourg, P. (2007). The effects of explicit referencing in distance problem solving over shared maps. In *Proceedings of the 2007 International ACM Conference on Supporting Group Work (GROUP '07)*, 4–7 November 2007, Sanibel Island, FL, USA (pp. 331–340). New York: ACM. <http://dx.doi.org/10.1145/1316624.1316674>
- Coles, S. (2001). *An introduction to statistical modeling of extreme values*. London: Springer. <http://dx.doi.org/10.1007/978-1-4471-3675-0>

(2017). An application of extreme value theory to learning analytics: Predicting collaboration outcome from eye-tracking data. *Journal of Learning Analytics*, 4(3), 140–164. <http://dx.doi.org/10.18608/jla.2017.43.8>

- Duchowski, A. T., Cournia, N., Cumming, B., McCallum, D., Gramopadhye, A., Greenstein, J., ... Tyrrell, R. A. (2004). Visual deictic reference in a collaborative virtual environment. In *Proceedings of the 2004 Symposium on Eye Tracking Research & Applications (ETRA '04)* 22–24 March 2004, San Antonio, TX, USA (pp. 35–40). New York: ACM. <http://dx.doi.org/10.1145/968363.968369>
- Grant, E. R., & Spivey, M. J. (2003). Eye movements and problem solving guiding attention guides thought. *Psychological Science*, 14(5), 462–466. <http://dx.doi.org/10.1111/1467-9280.02454>
- Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, 11(4), 274–279. <http://dx.doi.org/10.1111/1467-9280.00255>
- Jermann, P., Mullins, D., Nüssli, M. A., & Dillenbourg, P. (2011). Collaborative gaze footprints: Correlates of interaction quality. In *Connecting Computer-Supported Collaborative Learning to Policy and Practice: Proceedings of the 9<sup>th</sup> International Conference on Computer-Supported Collaborative Learning (CSCL 2011)*, 4–8 July 2011, Hong Kong, China (Vol. 1, No. EPFL-CONF-170043, pp. 184–191). International Society of the Learning Sciences.
- Jermann, P., & Nüssli, M.-A. (2012). Effects of sharing text selections on gaze cross-recurrence and interaction quality in a pair programming task. In *Proceedings of the 2012 ACM Conference on Computer Supported Cooperative Work (CSCW '12)*, 11–15 February, Seattle, WA, USA (pp. 1125–1134). New York: ACM. <http://dx.doi.org/10.1145/2145204.2145371>
- Jermann, P., Nüssli, M.-A., & Li, W. (2010). Using dual eye-tracking to unveil coordination and expertise in collaborative Tetris. In *Proceedings of the 24th BCS Interaction Specialist Group Conference (BCS '10)*, 6–10 September 2010, Dundee, UK (pp. 36–44). Swindon, UK: BCS Learning & Development Ltd.
- Kahrmanis, G., Chounta, I. A., & Avouris, N. (2010). Study of correlations between logfile-based metrics of interaction and the quality of synchronous collaboration. *International Reports on Socio-Informatics*, 7(1), 24–31.
- McNeil, A., Frey, R., & Embrechts, P. (2015). *Quantitative risk management: Concepts, techniques and tools*. Princeton, NJ: Princeton University Press.
- Meier, A., Spada, H., & Rummel, N. (2007). A rating scheme for assessing the quality of computer-supported collaboration processes. *International Journal of Computer-Supported Collaborative Learning*, 2(1), 63–86. <http://dx.doi.org/10.1007/s11412-006-9005-x>
- Meyer, A. S., Sleiderink, A. M., & Levelt, W. J. (1998). Viewing and naming objects: Eye movements during noun phrase production. *Cognition*, 66(2), B25–B33. [http://dx.doi.org/10.1016/S0010-0277\(98\)00009-2](http://dx.doi.org/10.1016/S0010-0277(98)00009-2)
- Nüssli, M.-A. (2011). Dual eye-tracking methods for the study of remote collaborative problem solving. PhD Thesis, École Polytechnique Fédérale de Lausanne.
- Reingold, E. M., Charness, N., Pomplun, M., & Stampe, D. M. (2001). Visual span in expert chess players: Evidence from eye movements. *Psychological Science*, 12(1), 48–55. <http://dx.doi.org/10.1111/1467-9280.00309>
- Richardson, D. C., Dale, R., & Kirkham, N. Z. (2007). The art of conversation is coordination common ground and the coupling of eye movements during dialogue. *Psychological Science*, 18(5), 407–413. <http://dx.doi.org/10.1111/j.1467-9280.2007.01914.x>

(2017). An application of extreme value theory to learning analytics: Predicting collaboration outcome from eye-tracking data. *Journal of Learning Analytics*, 4(3), 140–164. <http://dx.doi.org/10.18608/jla.2017.43.8>

- Richardson, D. C., Dale, R., & Tomlinson, T. M. (2009). Conversation, gaze coordination, and beliefs about visual context. *Cognitive Science*, 33(8), 1468–1482. <http://dx.doi.org/10.1111/j.1551-6709.2009.01057.x>
- Ripoll, H., Kerlirzin, Y., Stein, J.-F., & Reine, B. (1995). Analysis of information processing, decision making, and visual strategies in complex problem solving sport situations. *Human Movement Science*, 14(3), 325–349. [http://dx.doi.org/10.1016/0167-9457\(95\)00019-0](http://dx.doi.org/10.1016/0167-9457(95)00019-0)
- Schneider, B., & Blikstein, P. (2015). Comparing the benefits of a tangible user interface and contrasting cases as a preparation for future learning. In *Proceedings of the 11<sup>th</sup> International Conference on Computer Supported Collaborative Learning (CSCL 2015)*, 7–11 June 2015, Gothenburg, Sweden. International Society of the Learning Sciences.
- Sharma, K., Caballero, D., Verma, H., Jermann, P., & Dillenbourg, P. (2015). Looking AT versus looking THROUGH: A dual eye-tracking study in MOOC context. *Proceedings of the 11<sup>th</sup> International Conference on Computer Supported Collaborative Learning (CSCL 2015)*, 7–11 June 2015, Gothenburg, Sweden. International Society of the Learning Sciences.
- Sharma, K., Chavez-Demoulin, V., & Dillenbourg, P. (2017). Non-stationary modeling of tail-dependence of two subjects' concentration. (To appear) *Annals of Applied Statistics*.
- Sharma, K., D'Angelo, S., Gergle, D., & Dillenbourg, P. (2016). Visual augmentation of deictic gestures in MOOC videos. In *Proceedings of the 12<sup>th</sup> International Conference of the Learning Sciences (ICLS '16)*, 20–24 June 2016, Singapore. ISLS. <http://dx.doi.org/10.22318/icls2016.28>
- Sharma, K., Jermann, P., Nüssli, M.-A., & Dillenbourg, P. (2012). Gaze evidence for different activities in program understanding. In *Proceedings of the 24th Annual Conference of Psychology of Programming Interest Group*. London, UK, November 21–23, 2012. <https://infoscience.epfl.ch/record/184006>
- Sharma, K., Jermann, P., Nüssli, M.-A., & Dillenbourg, P. (2013). Understanding collaborative program comprehension: Interlacing gaze and dialogues. In *Proceedings of the 10<sup>th</sup> International Conference on Computer-Supported Collaborative Learning (CSCL 2013)*, 15–19 June 2013, Madison, WI, USA. International Society of the Learning Sciences. <https://infoscience.epfl.ch/record/184007>
- Stein, R., & Brennan, S. E. (2004). Another person's eye gaze as a cue in solving programming problems. In *Proceedings of the 6th International Conference on Multimodal Interfaces (ICMI '04)* 13–15 October 2004, State College, PA, USA (pp. 9–15). New York: ACM. <http://dx.doi.org/10.1145/1027933.1027936>
- Thomas, L. E., & Lleras, A. (2007). Moving eyes and moving thought: On the spatial compatibility between eye movements and cognition. *Psychonomic Bulletin & Review*, 14(4), 663–668. <http://dx.doi.org/10.3758/BF03196818>
- Van Gog, T., Jarodzka, H., Scheiter, K., Gerjets, P., & Paas, F. (2009). Attention guidance during example study via the model's eye movements. *Computers in Human Behavior*, 25(3), 785–791. <http://dx.doi.org/10.1016/j.chb.2009.02.007>
- Van Gog, T., Kester, L., Nieveelstein, F., Giesbers, B., & Paas, F. (2009). Uncovering cognitive processes: Different techniques that can contribute to cognitive load research and instruction. *Computers*

(2017). An application of extreme value theory to learning analytics: Predicting collaboration outcome from eye-tracking data. *Journal of Learning Analytics*, 4(3), 140–164. <http://dx.doi.org/10.18608/jla.2017.43.8>

*in Human Behavior*, 25, 325–331. <http://dx.doi.org/10.1016/j.chb.2008.12.021>

Van Gog, T., & Scheiter, K. (2010). Eye tracking as a tool to study and enhance multimedia learning. *Learning and Instruction*, 20(2), 95–99. <http://dx.doi.org/10.1016/j.learninstruc.2009.02.009>

Wise, A., & Shaffer, D. W. (2015). Why theory matters more than ever in the age of big data. *Journal of Learning Analytics*, 2(2), 5–13. <http://dx.doi.org/10.18608/jla.2015.22.2>

Zelinsky, G. J., & Murphy, G. L. (2000). Synchronizing visual and language processing: An effect of object name length on eye movements. *Psychological Science*, 11(2), 125–131. <http://dx.doi.org/10.1111/1467-9280.00227>