

# An Alternate Statistical Lens to Look at Collaboration Data: Extreme Value Theory

Kshitij Sharma, Department of Computer Science, Norwegian University of Science and Technology,  
Kshitij.sharma@ntnu.no

Jennifer Olsen, Computer Human Interaction Lab in Learning and Instruction, École Polytechnique Fédérale de  
Lausanne, Jennifer.olsen@epfl.ch

**Abstract:** To provide beneficial feedback to students during their collaboration, it is important to identify behaviors that are indicative of good collaboration. However, in a collaborative learning session, students engage in a range of behaviors and it can be difficult to indicate which of those behaviors correlate with higher outcomes. In this paper, we propose using Extreme Value Theory (EVT), a method that considers the data points in the tail (upper or lower) of the distribution, to analyse the relationship between collaborative process variables and outcome measures through insights derived from high impact, low-frequency events. Specifically, in this paper, we analyse the relationship between dual gaze patterns and outcome measures across two different datasets. In both datasets we found that students with lower outcomes had lower focus during the collaborative session. This paper provides a contribution by both introducing EVT as a viable method for analysing CSCL data as well as demonstrating the effectiveness of eye-tracking as a collaborative indicator to use to adapt to in real-time.

**Keywords:** Eye-tracking, dual eye-tracking, extreme value theory, CSCL, collaborative process, intelligent tutoring systems, concept maps.

## Introduction

Collaboration can be an effective way of learning, but it is challenging to ascertain how students' actions lead to learning when working in a group and how to determine what actions should be considered when providing students with feedback on their collaboration. To provide timely and effective feedback, it is important to have an indicator of collaboration that can easily be measured in real-time and is a predictor of learning. However, across collaboration indicators, the connections between these student process measures and learning outcomes have not always been clear (Deiglmayr, Rummel, & Loibl, 2015; Olsen, Aleven, & Rummel, 2017). In this paper, we propose analysing only the high impact, low frequency data points to provide additional insights into the relationship between process data and collaborative learning outcomes. We introduce the use of Extreme Value Theory (EVT), a method that considers the data points in the tail (upper or lower) of the distribution, to analyse the relationship between collaborative process variables and outcome measures. Specifically, we are interested in dual eye-tracking as a process variable as it can easily be measured in real-time and be used to provide students with feedback (Sharma et.al, 2016, D'Angelo & Begel, 2017). Using EVT, we aim to increase our understanding of the process variables that impact student collaborations and how they can be used to provide students with real-time feedback.

EVT is a novel method to CSCL and presents a complementary viewpoint to the statistical methods often used to analyse collaborative learning data while having fewer data assumptions. EVT can be used for both explanatory analyses (Ramesh and Davison, 2002) and hypothesis verification (Santinelli et.al., 2014). The mathematical foundations of EVT are similar in strength, rigor and maturity as compared to the methods based on the central tendencies in the data (Smith 1990). EVT provides a way to estimate the probability of occurrence of rare events, which might also be unseen in the observed data. This makes EVT a unique method to implement proactive feedback to support collaboration. For example, in the context of collaborative eye-tracking, if we could estimate the probability of peers not focusing on any part of the communication mediating interface, we could suggest some remedial actions in a proactive manner. In contexts where the proactive feedback is required, the tail of the data may contain more information about the process than the main body of the distribution.

In this paper, we apply EVT to eye-tracking as it has become a key source of process data in educational research over the past few years. Research using eye-tracking covers a wide range of educational ecosystems from online (Sharma et. al., 2015) to face to face classes (Raca & Dillenbourg, 2013), from co-located (Schneider et. al, 2016) to remote collaborative learning (Sharma et. al., 2012), and to understand teachers' classroom orchestration processes (Prieto et. al., 2016). Eye-tracking has not only been used to understand the learning processes in various contexts, but it also has been used to provide students appropriate, real-time, and adaptive feedback on their learning processes (Sharma et.al, 2016, D'Angelo & Begel, 2017).

One of the most common practical uses of dual eye-tracking (DUET) data in CSCL has been to quantify collaborative outcome and processes. In terms of collaborative outcome, recent results have shown DUET measures of gaze cross-recurrence (looking at the same area of the screen at the same time) to be useful in quantifying collaboration quality (Schneider et. al., 2013; Jermann and Nuessli, 2012) and the differences between the expertise (Papavlasopoulou et. al., 2017). Further looking at the similar areas of the screen in each time window was found to be correlated with learning gains (Sharma et. al., 2015; Schneider et. al., 2016; Sangin et. al., 2011). DUET measures such as average entropy (Sharma et. al., 2012), transitions among areas of interest (Villamor and Rodrigo, 2018) were correlated with task based performance in pair programming tasks. In terms of collaborative processes, eye-tracking research has shown to be useful in relating gaze to dialogue patterns such as the eye-voice span (time difference between the speaker's gaze at an object and verbalization, Allopenna et al., 1998), voice-eye span (time difference between the speaker's voice and listener's gaze to the referred object, Griffin and Bock, 2000) and eye-eye span (time difference between the speaker's gaze at an object and the listener's gaze at the same object, Richardson et al., 2007). Gaze cross-recurrence was also useful in explaining the conceptual knowledge gains of the peers in a collaborative learning scenario with intelligent tutoring systems (Belenky et. al., 2014). One common theme across these studies are the statistical methods used. Most of these studies used methods based on the central tendencies of the data. We propose a complementary viewpoint for the DUET data in this contribution that is based on the extreme values present in the data.

Within this paper, we aim to answer the research question of how the visual focus of peers is related to the collaborative learning outcomes. We propose two types of shifts from the traditional analyses. First, we move from the main body of the distribution (central tendency) to the tails of the distribution (extremes). Second, we move from the individuals (or groups) as a unit of analyses to the specific moments in the interaction. Specifically, we are interested in *how the extreme values (i.e., the moments with extremely high or extremely low values) can inform us of the students' learning gains*. To answer this question, we present the EVT method in two different dual eye-tracking (DUET) contexts. The first study is in a collaborative concept map context with university students, and the second study is in a collaborative Intelligent Tutoring Systems (ITS) context with elementary school children. For both contexts, we hypothesized that the gaze of the students with lower test scores will have a higher tendency to wander all over the screen. In the following sections, we will present the EVT methodology and how it can be used to provide insights into effective student collaboration processes. This paper contributes to the CSCL literature by providing an alternate method for analysing relationships between process and outcome data that complements existing methodology while also extending our understanding of the relationship between visual focus and collaborative outcomes.

## Method

### Extreme Value Theory

Extreme events are those with high impact and low frequency (HILF), and EVT is the branch of statistics used for modelling HILF events. The basic idea is to model an extreme event in a way such that the analyses and implications account for the needs of unprecedented situations. For example, in finance and environmental sciences, HILF events are often analysed to allow people and companies to be proactive against negative events (e.g., risk, losses, natural disasters). We can extend this proactive policy to education by considering negative and positive HILF events. EVT deals with asymptotic data where the data used in the analyses is a small subset of the whole data -- usually below 5th percentile or above 95th percentile. Mathematically, EVT is based on the tail of a given distribution. Because EVT is based on the tails, it does not impose any assumption on the distribution and can be applied to any known (e.g., normal, student, uniform, exponential) or unknown distribution. EVT was developed initially for independent and identically distributed variables (for a complete introduction to Extreme Value Theory, see Coles, 2001), but it can easily be extended to stationary variables (variables whose distributions are not affected by the time shifts), which are most variables addressed in the CSCL community.

### Advantages and disadvantages

There are three main advantages of EVT. **First**, since EVT does not have any assumptions for the data distribution, it can be used to analyse the tail (data below the 5th percentile or above the 95th percentile) of any given distribution. This makes EVT applicable with those datasets which cannot be analysed using the common parametric approaches. For example, ANOVA requires the data to follow a normal distribution and if not, normalisation operations can affect the interpretability of results. **Second**, when analysing the dependence structure between two variables, such as correlations, EVT does not assume the dependency to follow a known structure. For example, in case of correlations, this structure is assumed to be linear. This lack of assumption allows EVT to be applied irrespective of the nature of the distribution that generates the data and dictates the

relationships among different variables. The **third** advantage of EVT is over non-parametric models. Non-parametric models, which are used in CSCL as a way of hypothesis testing, provide one value (p-value). These methods summarise the data and can handle only low dimensional problems. Given the advents in big data, non-parametric, which are designed for smaller datasets, are at an inherent disadvantage. Also in time series analyses, dynamic models provide much more information about the process than non-parametric models. EVT accounts for time series analyses by providing methods to consider the covariate dependence on time while analysing CSCL data.

The main disadvantage of using EVT is related to the fact that EVT considers the tail of the data only. By only considering the tail, the parameter estimation becomes more difficult for datasets with few events. In other words, EVT is appropriate to be used contexts where the data is sufficiently large. However, in the “big data era” this problem is often solved by itself and in these circumstances the methods based on EVT provide robust approach for estimating the probabilities of rare events.

## EVT: Univariate

There are two ways to model data using EVT: blockwise maxima and points over threshold. In **blockwise maxima**, we divide the set of observations into  $M$  blocks of  $N$  data points. This results in a sequence of maximum values for each block. One of the most important results of EVT formulation is that for any given distribution for the main body of data, known or unknown, the sequence of maximas follow a unique distribution called Generalised Extreme Value (GEV). Once we have the sequence of maximas, we simply fit the GEV distribution on them using any likelihood optimisation technique. This distribution is characterized by three parameters: position, scale and shape. These values are then used to estimate the value which has a low probability of being exceeded (see further in this section).

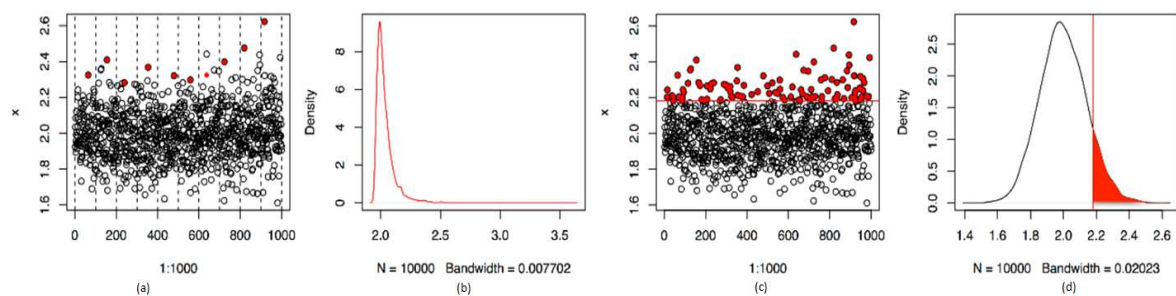
Using **points over threshold method**, we consider all the data points over a threshold. This threshold is usually a high percentile of the observed data (90th or 95th percentile). These data points above the threshold are called the exceedances. Using points over threshold method, the exceedances follow a Poisson distribution and the exceedance size follows a Generalised Pareto Distribution (GPD). The parameters of GPD can be estimated by GEV. The shape parameter for both GPD and GEV remains the same. This parameter determines how long/heavy the tail of the observed data is. Once we have the threshold, the rate parameter of the Poisson distribution and the scale and shape parameters of GPD can be estimated, like GEV, by using a likelihood optimization technique. Figure 1 shows the difference in the two approaches.

The position (or rate of Poisson), scale and shape parameters are difficult to interpret in CSCL situations, since these values do not correspond to any behavioural indicator. Therefore, we calculate a quantile value at a high level (above 90th or 95th percentile), which has an important interpretation. This value is called **Return Level**, which represents a measure of an extreme event, possibly unseen, with a certain probability. For example, if a return level is calculated at the 95th percentile, this indicates that the actual (unseen) extreme event will exceed this value with a 0.05 probability. In other words, a return value at 95th percentile is the value that has a probability of 0.05 of being exceeded. One can estimate this value by position, shape and scale measures from GEV or rate of arrival of exceedances, shape and scale measures of GPD.

One might inquire that if the high percentile can be directly calculated from the data, then why we need this estimation. The answer to this question is that we could surely calculate the return levels from the data, however minor discrepancies in the data would result in large errors for such computation, resulting in an erroneous return value larger or smaller than expected. If the return value is larger than it should be then one cannot proactively take actions; and if the return value is smaller, then one will be acting upon at a wrong time. It can be shown, mathematically that estimating GEV or GPD parameters are the correct way of computing the return level in the data. Finally, once we have the return level for the collaborating partners, we can compute the difference between them: by value and/or when they appear, and this difference can be used as a measure to further correlate against the quality/outcome of collaboration.

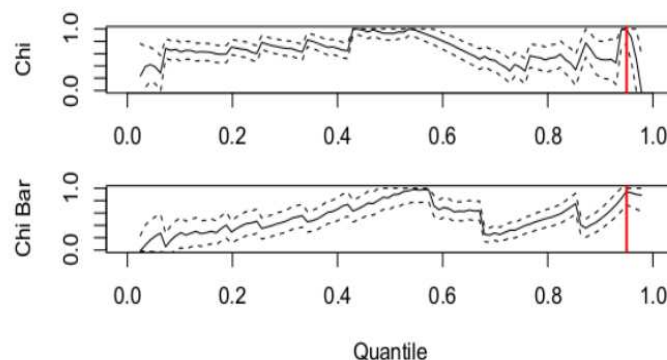
## EVT: Bivariate

To analyse the time series data from peers, the bivariate case of EVT can be useful. Bivariate EVT measures the extremal dependence between two time series data, such as that between a collaborating dyad. It models the probability of observing an extreme event in one time series given that there is an observable extreme event in the other one. This probability can be quantified using the tail-dependence between the two time series. In a classical statistical approach, the dependence between two time series is measured by correlation, which is computed using the central tendencies of the data. In the case of EVT, the tail-dependence is calculated at the high percentile of the data, as in case with the return levels.



**Figure 1.** (a) A random variable simulation with the blockwise-maxima; (b) density plots for one of the blocks; (c) the red horizontal line shows the threshold for the POT method; (d) the whole distribution, the red vertical line shows the threshold for the POT method and denotes the beginning of the tail for the distribution.

There are two measures of extremal dependence, originating from classical multivariate EVT: asymptotic dependence and asymptotic independence. The coefficient of asymptotic dependence (CAD) is the tendency for one variable to be over a high threshold when the other exceeds this threshold. This value is always between 0 and 1. The only possibility of asymptotic independence is when CAD is 0. When the CAD is greater than 0, the variables are asymptotically dependent. On the other hand, the coefficient of asymptotic independence (CAI) is the measure of strength of this extremal dependence. This is measured by a conditional probability that the smaller values in the time series of one variable are below a infinitesimal threshold, given the smaller values in the other time series are below that threshold. Mathematically, it can be shown that a value of 1 for CAI shows the perfect dependence and a value of 0 for CAI shows perfect independence. In summary, the CAD shows the level of asymptotic dependence between two time series while the CAI shows the strength of this dependence. Figure 2 shows an example of how to determine CAD and CAI.



**Figure 2.** Example illustrating the determination of the coefficient of extremal dependence and the strength of dependence for a typical time-series pair. The dashed lines represent the 95% confidence intervals for CAD (top, Chi) and CAI (bottom, Chi Bar). 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%.

## Examples

In this section, we provide two different examples from distinct collaborative learning scenarios. For each context, we show how EVT can be applied on one gaze variable computed from both the studies, and how the different EVT based measurements differentiate the quality/performance levels in collaborative learning outcomes.

### Collaborative concept map

This data set involves 24 dyads from a larger study that tested a hypothesis about the relation between individual and collaborative gaze patterns (Sharma et. al., 2015). Each dyad was engaged in a collaborative concept-map building activity. The students were sitting on two sides of a visual separation and could talk to each other. Prior to the concept-map building activity, they individually watched two videos from Khan Academy about resting membrane potential and they were asked to build the concept-map about the same topic. The main task was to relate the pre-defined concepts and add new concepts if they felt necessary. The students watched the video at their own pace and the total duration for the concept-map activity for each dyad was between 10 and 12 minutes. For this contribution, the dependent measure is calculated as follows. 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.

### Collaborative intelligent tutoring systems

This data set involves 14 4th and 14 5th grade dyads from a larger study that tested the hypothesis about differential benefits of collaborative versus individual learning (Olsen et. al., 2014). The dyads were engaged in a problem-solving activity around fractions using a networked collaborative ITS, which allowed them to synchronously work in a shared problem space where they could see each other's actions while sitting at their own computers across the room from each other. The students could communicate verbally through a Skype connection. Each dyad worked with the tutor for 45 minutes in a pull-out study design at their school. The morning before working with the tutor and the morning after working with the tutor, students were given 25 minutes to complete a pretest or posttest individually on the computer to assess their learning. During the experiment, dual eye tracking data, dialogue data, and tutor log data in addition to the pretest and posttest measures were collected. For this contribution, the dependent measure is the average posttest score of each dyad.

### Variable: Spatial Entropy

To capture the visual focus, we use Spatial Entropy (SE) that is one of the measures used to analyse DUET data in previous research (Olsen et. al., 2018; Sharma et. al., 2018) and show the results base on EVT analyses for both learning contexts. SE measures the spatial distribution of the gaze of each peer. To compute SE, we first define a 50-pixel by-50-pixel grid over the screen and we compute for each peer the proportion of fixation time located in each grid cell (Figure 3). This proportion is computed over a time window of five seconds. 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.

What does extreme spatial entropy mean? The idea is to capture the visual focus size (not attention, although in the contexts of the two examples they might be related) of the participants. The higher the spatial entropy is, the larger the focus size is. A spatial entropy value of zero indicates that the participant is looking at only one part of the screen and higher values indicate that the participant looks at different parts of the screen, during a given time window. Now, an extreme spatial entropy would indicate that the participant is looking “*all over the place*”.

## Results

### Collaborative concept map

First, using the univariate EVT, we check the average return levels of the two spatial entropy time series. The average return levels (calculated at 95<sup>th</sup> percentile) for spatial entropy is lower for the pairs with high collaboration outcome ( $F[1,15.78] = 6.53$ ,  $p\text{-value} = .01$ , one-way ANOVA without assuming equal variances) than the average return levels of spatial entropy for the pairs with low collaboration outcome.

Second, considering bivariate EVT results for the spatial entropy of the peers, there are two values to be checked: 1) the level of extremal dependence and 2) the strength of extremal dependence if the level is non-zero. We observe a higher extremal dependence (calculated at the 95% quantile) between the spatial entropy of peers with low collaboration outcome ( $F[1,22] = 4.28$ ,  $p\text{-value} = 0.01$ ) than the extremal dependence between the spatial entropy of peers with high collaboration quality. We observe an even more significant difference in the strength of extremal dependence (calculated at the 95% quantile) for the pairs with the two different levels of collaboration outcome ( $F[1,22] = 10.43$ ,  $p\text{-value} = 0.001$ ). Pairs with low level of collaboration outcome have stronger extremal dependence between the spatial entropy of peers than that for the pairs with high level of collaboration outcome.

### Collaborative intelligent tutoring systems

In the univariate EVT case for the ITS data, the average return levels (calculated at 95<sup>th</sup> percentile) for spatial entropy is negatively correlated with the pair's average posttest score ( $\text{cor} = -0.48$ ,  $p = .01$ ). Next, we look at the bivariate EVT for the ITS data. There is a negative correlation between the upper tail dependence (calculated at the 95% quantile) for the spatial entropy of peers and the average post test score of the pairs ( $\text{cor} = -0.48$ ,  $p = .01$ ). In the case of ITS data, we observe a significant and negative correlation between the strength of upper tail

dependence (calculated at the 95% quantile) for the spatial entropy or pairs and their average posttest score ( $\text{cor} = -0.43, p = .03$ ).

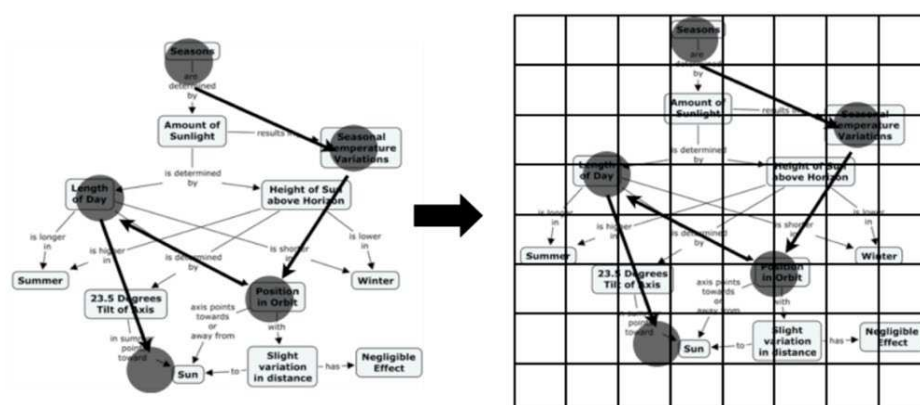


Figure 3. The process of computing entropy. The image on the left shows the exemplar concept-map and gaze patterns (grey circles and arrows). The image on the right shows the placement of the grid.

## Discussion

In this paper, we presented a new method in the context of analysing CSCL data, specifically dual eye-tracking data. We propose that EVT based methods are robust enough to estimate the probabilities of the rare events, which can then be used to provide proactive feedback to students. As examples, we presented results from two dual eye-tracking studies: collaborative concept map and collaborative ITS. To explain the findings from both the studies in a unified way, we use the same metric to capture the focus (spatial entropy) of the peers in two studies.

In the univariate EVT case, we propose to use the return value at the 95th percentile. This value of the spatial entropy has 5% chances of being exceeded. Across both the concept map and ITS contexts, the results indicate that the average values of extreme entropy is higher for the pairs with the low collaborative outcome/learning than the pairs with high collaborative outcome/learning. This simply translates to the fact that pairs with high levels of collaborative outcome/learning have lower levels of spread-out gaze patterns.

The bivariate context provides a supporting explanation for the results from the univariate EVT. We observe that both the level and strength for the upper tail distribution of the spatial entropy is negatively correlated with the collaborative performance/learning, i.e., the moments of extreme entropy appear together in time for the pairs with low collaborative outcome/learning. Combining this with the univariate results, we can conclude that the peers with low collaborative outcome/learning not only have higher chances of “looking all over the place” but there are high chances of them looking all over the place at the same time.

This indicates that the pairs with low collaborative outcome have moments where both the participants in the pair have extremely large visual focus. The cause of such behavior could be explained in two different ways. First, both the participants are looking for some information on the screen and thus they have a large visual focus size. Second, the mutual understanding has a missing link that needs to be created between the two peers. This information can be used to intervene proactively by providing greater scaffolding for student interactions to help guide the collaborative learning process when students have low focus at the same time.

In terms of contemporary dual eye-tracking analyses methods, the bivariate EVT is like analysing the gaze cross-recurrence. Gaze cross recurrence (CR) has been found to be correlated with the collaboration quality (Jermann and Nuessli, 2012; Schneider et.al, 2013). CR indicates the time peers spent looking at the same area on the screen at the same time. In the bivariate EVT case, visual focus of the peers is compared over time. Having comparable visual focus size does not necessarily mean that the peers are looking at the same part of the screen. However, having a large focus size in each time window will result in a high CR over time, as in an aggregated time frame the peers would be looking at the same area on the screen, which is the whole screen. One interesting finding from the examples presented in this paper is the relation between the extreme visual focus and the average learning gains of the peers. In a recent contribution (Olsen, Aleven, & Rummel, 2017), the researchers did not find a relation between CR and overall learning gains of the students using the same collaborative ITS data. This suggest that by analysing the HILF data, we may be able to gain additional insights that are not apparent when analysing the entire data set.

The relation between the extreme visual focus and collaborative outcome/learning, provides an opportunity for designing proactive feedback tools to support collaboration. Using the methods described in this

contribution, one can identify the key moments to provide the feedback to the collaborators. Most of the recent work done in the direction of using gaze awareness to scaffold collaboration has been focused onto showing the gaze of the peers to each other. For example, Ishii and Kobayashi (1992) and Monk and Gale (2002) designed systems displaying the face of the collaborators. Stan and Brennen (2004) showed that displaying partners' gaze while debugging a program helped finding bugs. In another experiment, Brennan, et. al. (2008) showed that displaying partners' gaze in "Os-in-Q" search helped the collaborators in more effective labor division. Recently, the gaze-awareness has been used to support collaboration in high level tasks such as pair programming (D'Angelo and Begel, 2017). Gaze awareness has also been shown to be useful in co-located collaboration scenarios (Van Rheden et. al., 2017) and remote collaborations using different types of devices (Akkil, et, al., 2018). One common theme across these gaze visualisations is that the gaze is visualized throughout the whole collaborative work, which might hinder the learning experience on a few occasions. Identification of the key moments during the collaboration for providing support might improve the effectiveness of the gaze aware tools.

With the results from the two DUET studies based on the EVT, we can estimate a high value of visual focus for both the participants which has a very low probability of being exceeded. In a collaborative scenario if we observe that the visual focus sizes for both the participants in the dyad is going to exceed a certain threshold, we can proactively support the pair to avoid disruptions in the collaboration. EVT has these added values for gaze-aware feedback systems, which could enable the instructor to be proactive and select the exact moments to provide guidance rather than visualizing the support throughout the collaboration.

## Conclusions

We presented a complementary method to analyse CSCL data based on the tails (lower or upper) of the distribution. This method is based on a mathematical theory, which is novel to CSCL community, called Extreme value theory. The main motivation behind using EVT is to be able to provide proactive scaffolding during the moments, when the collaboration between peers is at a point where the outcome/quality is disruptive. We do not claim the superiority of this method (for a comparison with the traditional methods of analyses, see Sharma et. al., 2016). We propose, that EVT provides a different point of view for the data when the traditional methods do not apply because of the failed assumptions or when the traditional methods do not provide any useful insights about the collaborative processes, outcome or quality.

Within this paper, we provide a contribution both through the application of EVT to educational process data as well as furthering the understanding of how dual eye-tracking relates to collaborative learning outcome measures. By analysing just, the tail of the data, we can distinguish patterns that may not have otherwise been apparent. Across contexts, we have shown a common pattern of with low outcome measures having lower focus at the same time. We propose EVT as an alternative method for the analysis of collaborative learning data that can provide complementary viewpoints to the common CSCL methodological repertoire.

## References

- Akkil, D., & Isokoski, P. (2018). I see what you see: gaze awareness in mobile video collaboration. In Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications (p. 32). ACM.
- Allopenna, P.D., Magnuson, J.S., and Tanenhaus, M.K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models\* 1,\* 2,\* 3,\* 4,\* 5. *Journal of memory and language*, 38(4),
- Bednarik, R., & Kauppinen, M. (2013). Unravelling the interaction strategies and gaze in collaborative learning with online video lectures. In Proceedings of the 6th workshop on Eye gaze in intelligent human machine interaction: gaze in multimodal interaction (pp. 57-62). ACM.
- Belenky, D., Ringenberg, M., Olsen, J., Alevan, V., & Rummel, N. (2014). Using Dual Eye-Tracking to Evaluate Students' Collaboration with an Intelligent Tutoring System for Elementary-Level Fractions. Grantee Submission.
- Coles, S. (2001). *An introduction to statistical modeling of extreme values*. London: Springer. <http://dx.doi.org/10.1007/978-1-4471-3675-0>
- D'Angelo, S., & Begel, A. (2017). Improving communication between pair programmers using shared gaze awareness. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (pp. 6245-6290). ACM.
- Deiglmayr, A., Rummel, N., & Loibl, K. (2015). The Mediating Role of Interactive Learning Activities in CSCL: An Input-Process-Outcome Model In Lindwall, O., Häkkinen, P., Koschman, T. Tchounikine, P. Ludvigsen, S. (Eds.) (2015). *Exploring the Material Conditions of Learning: The Computer Supported Collaborative Learning (CSCL) Conference 2015, Volume 2*. Gothenburg, Sweden: The International Society of the Learning Sciences.

- Embrechts, P. (2000). Extreme Value Theory: Potentials and Limitations as an Integrated Risk Management Tool. Manuscript, Zurich, Switzerland: Department of Mathematics, ETH, Swiss Federal Technical University.
- Fussell, S. R., Setlock, L. D., & Parker, E. M. (2003). Where do helpers look?: gaze targets during collaborative physical tasks. In CHI'03 Extended Abstracts on Human Factors in Computing Systems (pp. 768-769). ACM.
- Griffin, Z.M., and Bock, K. (2000). What the eyes say about speaking. *Psychological science*, 11(4).
- Ishii, H. and Kobayashi, M. (1992). Clearboard: a seamless medium for shared drawing and conversation with eye contact. In *Proceedings of CHI*, ACM.
- 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 ACM 2012 conference on Computer Supported Cooperative Work* (pp. 1125-1134). ACM.
- Monk, A.F., and Gale, C. (2002). A look is worth a thousand words: full gaze awareness in video-mediated conversation. *Discourse Processes*, 33(3), 257–278.
- Olsen, J. K., Alevén, V., & Rummel, N. (2017). Exploring Dual Eye Tracking as a Tool to Assess Collaboration. In *Innovative Assessment of Collaboration* (pp. 157-172). Springer, Cham.
- Olsen, J., Sharma, K., Alevén, V., & Rummel, N. (2018). Combining Gaze, Dialogue, and Action from a Collaborative Intelligent Tutoring System to Inform Student Learning Processes. *International Society of the Learning Sciences, Inc.[ISLS]*.
- Papavlasopoulou, S., Sharma, K., Giannakos, M., & Jaccheri, L. (2017). Using Eye-Tracking to Unveil Differences Between Kids and Teens in Coding Activities. In *Proceedings of the 2017 Conference on Interaction Design and Children*(pp. 171-181). ACM.
- Prieto, L. P., Sharma, K., & Dillenbourg, P. (2015). Studying teacher orchestration load in technology-enhanced classrooms. In *Design for teaching and learning in a networked world* (pp. 268-281). Springer, Cham.
- Ramesh, N. I., & Davison, A. C. (2002). Local models for exploratory analysis of hydrological extremes. *Journal of Hydrology*, 256(1-2), 106-119.
- Richardson, D. C., Dale, R., and Kirkham, N. Z., (2007). The art of conversation is coordination. *Psychological Science*, 18(5).
- Raca, M., & Dillenbourg, P. (2013, April). System for assessing classroom attention. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (pp. 265-269). ACM.
- Santinelli, L., Morio, J., Dufour, G., & Jacquemart, D. (2014). On the sustainability of the extreme value theory for WCET estimation. In *OASIS-OpenAccess Series in Informatics* (Vol. 39). Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.
- Schneider, B., Abu-El-Haija, S., Reesman, J., & Pea, R. (2013). Toward collaboration sensing: applying network analysis techniques to collaborative eye-tracking data. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (pp. 107-111). ACM.
- Schneider, B., Sharma, K., Cuendet, S., Zufferey, G., Dillenbourg, P., & Pea, R. (2016). Using mobile eye-trackers to unpack the perceptual benefits of a tangible user interface for collaborative learning. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 23(6), 39.
- Sharma, K., Chavez-Demoulin, V., & Dillenbourg, P. (2016). An Application of Extreme Value Theory to Learning Analytics: Predicting Collaboration Outcome from Eye-tracking Data.
- Sharma, K., Alavi, H. S., Jermann, P., & Dillenbourg, P. (2016). A gaze-based learning analytics model: in-video visual feedback to improve learner's attention in MOOCs. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge* (pp. 417-421). ACM.
- Sharma, K., Jermann, P., Nüssli, M. A., & Dillenbourg, P. (2013). Understanding collaborative program comprehension: Interlacing gaze and dialogues. In *Proceedings of Computer Supported Collaborative Learning (CSCL 2013)* (Vol. 1, No. EPFL-CONF-184007, pp. 430-437).
- Sharma, K., Caballero, D., Verma, H., Jermann, P., & Dillenbourg, P. (2015). Looking AT versus looking THROUGH: A dual eye-tracking study in MOOC context. *International Society of the Learning Sciences, Inc.[ISLS]*.
- Sharma, K., Olsen, J. K., Alevén, V., & Rummel, N. (2018). Exploring Causality Within Collaborative Problem Solving Using Eye-Tracking. In *European Conference on Technology Enhanced Learning* (pp. 412-426). Springer, Cham.
- Smith, R. L. (1990). Extreme value theory. *Handbook of applicable mathematics*, 7, 437-471.
- Van Rheden, V., Maurer, B., Smit, D., Murer, M., & Tscheligi, M. (2017). LaserViz: Shared gaze in the Co-located physical world. In *Proceedings of the Eleventh International Conference on Tangible, Embedded, and Embodied Interaction* (pp. 191-196). ACM.