Thomas Hennig

Variability Analysis based on Multi-Objective Performance and Throw Acceleration in Dart-Game

Intelligent Cooperative Systems
Computational Intelligence

Variability Analysis based on Multi-Objective Performance and Throw Acceleration in Dart-Game

Master Thesis

Thomas Hennig

November 20, 2018

Supervisors: Prof. Dr. Sanaz Mostaghim,
Chair of Computational Intelligence

Prof. Dr. Frank Ohl,
OvGU, Leibniz Institute for Neurobiology

Advisors: M.Sc. Heiner Zille

Dipl.-Neurowiss. Lars T. Boenke

Thomas Hennig: *Variability Analysis based on Multi-Objective Performance and Throw Acceleration in Dart-Game*
Otto-von-Guericke Universität
Intelligent Cooperative Systems
Computational Intelligence
Magdeburg, 2018.

Abstract

Movements are part of everyday activities, need to be learned and can be improved regarding their movement goal. In related work, occurring variance in movements is mostly considered as noise. Whereas theories from cybernetics and dynamic system theory suggest that part of this variance supports achieving the movement goal.

The aim of this thesis was to explore the relation of variability between throw performance and throw physiology in throwing darts. Differences between players that perform better and players that perform worse should be identified. In Addition, differences between players that improve more and others that improve less should be identified. In related work, a single measure is used to quantify performance in throwing dart. In this thesis, multiple measures should be used to quantify performance. For analyzing the relation between performance and the way of throwing, 6 visual analyses were conducted. In each analysis, a good and a bad group based on performance measures were confronted so that differences in their throw pattern could be identified. To quantify performance, two performance-measures for sets of throws were defined to represent scatter and average deviation. To quantify the variance of throw movements, the variance of an acceleration model that was fitted to every throw was determined.

A triangular relation between our performance measures was found that is neither the effect of their definition nor an effect solely created by our method. Additional unique differences were identified by analyzing both performance measures together using the fronts from non-dominated sorting. For the majority of observed differences in the throw pattern, a better performance was related to a smaller variance in the throw pattern. For few observed differences throwing, a better performance was related to a higher variance in the throw pattern. For the differences in the development of parts of the throw pattern, clearly less observed differences were related to the development of variance.

Different types of variability relations are suggested by our results. The variance of these relations is compound when quantifying the variance in the way of throwing. Further work needs to be done to separate these compound variances.

Preface

First of all, I like to thank Lars T. Boenke, Heiner Zille, Prof. Dr. Frank Ohl, and Prof. Dr. Sanaz Mostaghim for giving me the opportunity to work on this topic. I thank them for all their time for our many meetings, for all their support and all their advice during the work for this topic. I thank them and Dominik Weikert for their preceding work and ideas.

Next, I want to thank Prof. Ross Cunnington, Tamara Powell et al. for providing us with the data of the experiment for our analyses and for explaining these in detail.

Furthermore, I like to thank Dr. Christoph Steup for sharing his know-how in throwing darts and his expertise on analyzing measured data, as well as the students in the work group of Prof. Dr. Mostaghim for their advice and ideas.

It the end, as this thesis is also the completion of my studies, I like to thank my partner and my family for their support throughout the years.

Contents

List of Figures

List of Tables

1 Introduction

When we learn a second language, we learn words, phrases, grammar, pronunciation and as part of this also often sounds that are specific to the new language. For people with German as the first language, this might be the th-sound when learning English. The other way around, for people with English as the first language who learn German, there are the umlauts. To create the th-sound, we need to coordinate tongue and the tip of the upper teeth to form a slit through which we press air with our breath to create a hissing sound. For the umlauts, we take the example of the sound for *ü*. We start by coordinating the jaw, tongue, and lips to continuously create the English sound for the letter *e*. Now, keeping the jaw and the tongue in this position, we purse our lips slowly until just a fingertip-sized opening is left. We see that we need to coordinate the movement of multiple parts of our body to create a specific new sound. To use these sounds in the language, we need to learn to create those sounds reliably and fast - so, others may understand us. As a baby, there are not just a few sounds that we need to learn. We need to learn to move mouth, jaw, tongue and vocal cords and to coordinate this with the breath for the many sounds of the language of our parents.

For many everyday actions, we need to learn movements. For communication, apart from talking, we also need to learn the movements for writing whether using a pen or using a keyboard. We need to learn the movements for walking, for grabbing. After an accident, we may need to learn to talk again. We may need to learn to walk using an artificial limb. In sports, we need to learn movements. For football, we need the moments to run and control a ball at the same time. In darts or curling, we need a movement to accelerate a missile to hit a distant target.

We know that we move due to forces that apply to limbs and tissue. We know that these forces originate from muscles that contract. We know that these muscles contract due to signals provoked in the brain. We know the structure of the brain, that certain brain areas correspond to certain body parts. However, we do not know how it works so that movements emerge. We do not know how the brain learns a new movement. We do not know how the brain changes to improve the movement. If we would know, we could tell what to do to be an expert and how to get there. We could choose strategies to learn faster.

A question that is related to this is whether we execute always the exact same movement. [Bernstein1967] observed that eventhough the movement appears to be very similar, there is always a variability in the movement which also covers alternative movements for the case that some movements are not feasible. For example, we can walk. Yet, the movement differs as we walk on a rough ground or a slippery groung, as we walk a slope up, down or on a flat ground.

Our intention for this thesis is to make first steps in the exploration of the variability in movements. Our goal is to explore for throwing darts the relation between performance and the way of throwing with a focus on variability. As we see in the related work in Chapter 2 that usually just a single objective or in other words performance measures is used, we want to include multiple performance measures in our analyses. The movement of throwing darts as an object of investigation has several advantages compared to other movements. We have a precisely defined goal with the dartboard. The movement itself affects comparatively few limbs - mainly arm and hand. In addition, there are few constraints to throwing darts. So, we do not need exceptional strength, speed, agility, endurance or intelligence. In the exploration of the dart throwing movement, we follow two **research questions**.

- What are the differences between players that show a high performance and players that show a low performance?
- What are differences between players that improve more in their performance than others?

To find answers to these questions, we structure this thesis into the following parts.

- In the next chapter, Chapter 2, we present the basis to investigate the movement of throwing darts.
- In Chapter 3, we present the dataset that is the basis for our analyses.
- In Chapter 4, we describe our developed method to analyze this dataset.

- In Chapter 5, we present the results of applying our method to the dataset.
- In Chapter 6, we discuss the results, the strengths and weaknesses of our method, as well as options for future work.

2 Background

In this chapter, we want to provide the knowledge that is required to understand our method. We will start by describing the game of darts - what it is and how it works. Based on this, we will introduce the topic of motor control which gives us theories to explain how we create throw movements in darts and which leads us to our expectations for the following chapters. Afterward, we continue with the topic of quantifying throw performance as well as throw physiology which will be necessary for analyses.

2.1 Game of Darts

Darts is a popular game and competitive sport. Participants throw repeatedly from a defined distance small missiles - called **dart** - at a static target. The target is a disk - the **dartboard** - which is attached in a defined height at a wall. The dartboard is subdivided into multiple fields. In the center, there is a circular field called **bull's eye**. Around it, the dartboard is divided into several **rings** as well as it is divided into several **sectors** of a circle which forms a multitude of **ring segments**. Figure 2.1 shows the dartboard that we used for this thesis. It has a diameter of about 43 cm and is subdivided into 9 rings and 12 equal sized sectors that form 108 ring segments. There are different variants of the game with different goals in which participants may compete. These variants have in common that a certain amount of points is assigned to each field which is gathered for every hit in the respective field. In a variant of the game for more advanced participants, participants may seek to hit a certain combination of fields in order to gather points for their personal scores with the goal to reach a certain sum without exceeding it. On the other hand, beginners may just try to hit the bull's eye receiving more points the closer their hit is to the bull's eye. The process of throwing a dart starts with an acceleration phase in order to reach an appropriate velocity and

Figure 2.1: The dartboard that we used in this thesis - The red circular field in the center is the bull's eye. Around it, the dartboart is subdivided by black and white rings, as well as 12 sectors of a circle into fields in the shape of ring segments.

movement direction so the dart may fly from the participant to the dartboard bridging the distance between both. When the dart reaches an appropriate velocity then the participant releases the darts into flight. In flight, the dart is not influenced by the participant anymore. Instead, for example, gravity is influencing the dart which was compensated by the participant before. The flight phase ends with the impact of the dart. It hits the dartboard or the surrounding wall reducing it velocity greatly. On hitting the dartboard, the dart usually gets stuck. Under the name **hit**, we refer to the impact position of a single throw. Though, we name hits of throws that missed the dartboard also **misses**. Regarding the throw process of participants, we observe that participants prepare a throw by reaching an initial state that usually extends the distance for acceleration. Assuming that they already hold the dart in one hand, they may move their hands up to the level of the shoulder and backward. Afterward, they start accelerating this hand. Doing so, the hand moves forward towards maximum extension. There are many ways to accelerate the dart. To mention just two, we observed players mainly rotating their forearm around the elbow joint creating a rotational movement while for others, we saw an ejecting movement by expanding forearm and upper arm rotating in elbow joint and shoulder joint. Since there is a maximum extension of the arm, there must be a phase of deceleration that reduces the velocity of hand and arm. Reaching the maximum extension means that in this movement at some point the bones of upper arm and forearm collide in the elbow joint. To prevent such a hard break, participants also cancel acceleration before the maximum extension is reached and decelerate actively to stop the arm. Releasing the darts between those phases of acceleration and deceleration is a matter of timing. While the arm is used to accelerate the dart, it is the hand that holds the dart and releases it into flight. On releasing too early, the dart may not have reached the necessary velocity. On releasing too late, the dart may already be to slow again due to deceleration. The throw of a single dart is a complex process. It requires that the participant coordinates the hold and release of a dart with the acceleration and deceleration of the arm in order to hit a distant target. In the following section, we will take a look at theories that try to explain how we control those complex motoric processes.

2.2 Theories of Motor Control

Motor control is the discipline in which scientists study how movements arise in humans that suffice certain movement goals (cf. [Edwards2011]). This includes the question for the role of cognitive processes. In darts, an example movement goal is to create repeatedly a movement to throw the dart and hit the bull's eye. With the concept of **performance** we qualify the fulfillment of movement goals in this context (cf. [Edwards2011]). Permanently hitting the bull's eye fulfills constantly and to a maximum extent the goal which represents the highest performance. Hits close to the bull's eye fulfill the goal more - showing a higher performance - than hits that miss the dartboard. Based on motor control, the discipline in which scientists study how movements are acquired and their performance improved is called motor learning (cf. [Edwards2011]). This discipline encompasses - for example - our research question of what changes occur in throwing dart together with improvements in performance. So in the game of darts, learning is connected to the improvement in performance. For a detailed overview of the development and branches in motor control and motor learning, we like to recommend [Edwards2011]. Here, we just want to provide the steppingstones that lead to the expectations for our analyses.

According to [Edwards2011], the older branch in motor control are closed-system theories or also called cognitive-based theories. Applying these, we consider mind and body to be separated and that the mind exclusively controls the body. At the beginning of a movement the mind is fed by sensory inputs which are then processed by the mind to extract information about body and environment. Using the available information and following the movement goal, the mind chooses a certain movement for which it plans a program to steer the body. Afterward, the program is executed by sending signals to muscles which evoke actions. A movement is the sum of these actions. Applying this to darts, participants start by observing the environment and the arm in the initial throwing position. From these sensory inputs, they extract information on the vector towards the desired field of the dartboard. With this information, they decide, e.g., for a rotational throw movement that creates a throw trajectory towards that field. The participants start to plan which muscles, joints, and bones are needed to create this motion. After the planning is done, the signals are sent towards the muscles that extend the forearm regarding the upper arm by contracting and pulling the forearm accelerating. This accelerates the arm as well as the dart that is held in the hand. On another signal to the muscles of the hand, the dart is released into flight. On the last signal, the muscles of the arm stop accelerating the arm and start decelerating it instead until the arm rests again. The actions for acceleration, release, and deceleration form the movement of a throw.

[Bernstein1967] objected the idea of movements to be exclusively controlled by cognitive processes as he identified two problems that oppose the idea of cognitive-based theories. His first problem - the degree of freedom problem - describes that from limbs down to the level of muscles, joints, and bones, down to the level of muscle cells there are too many entities that would require control for a movement. So a controlling program for any movement would be too complex to be able to be efficiently executed. With his second problem - context-conditioned variability - he questioned the assumption of cognitive-based theories that the body is independent of external influences. He observed that depending on different external influences movements pursuing the same movement goal were generated differently. He observed that while slowly lowering a stretched arm, muscles in the back are pulling the arm down when another person is actively pushing the arm up. Whereas, when only gravity is affecting the arm then muscles in the shoulder are pulling the arm up.

Bernstein's chain of thoughts opposed an exclusive control of cognitive processes and showed a strong interrelation of body and environment in the generation movements. To explain how we generate movements despite the degrees of freedom, he introduced the model of temporary groupings of muscles, joints, and bones - called synergies - to certain action units that reduce the number of degrees of freedom. Doing so, he opened the alternative branch of dynamic-system theories in motor control in which we consider body, mind, and environment as complex, interrelated subsystems in which movements are emergent patterns from self-organizing structures in body and mind (cf. [Edwards2011]).

Another of Bernstein's observations was that even though movements that pursued the same movement goal were very similar, there were always variations. He named this phenomenon repetition without repetition as it describes repetition in performance without repetition of the precise movement. Based on this, [Latash2008] developed the concept of good and bad variance. While there is a variance in the movements that suffice the given goal there is another variance that reduces the performance. An example of good variance may be advanced darts players who show similar performance according to different techniques. They may execute a rotational throw and hit the bull's eye, they may execute an ejecting throw and hit the bull's eye, and they may throw backward above their shoulders and hit it again. On the other hand, we could introduce bad variance by starting to shake them during the throw which adds noise and degrades their performance.

From variance in motor control, we make a step to variety in cybernetics - "the science that studies the abstract principles of organization in complex systems" [Heylighen and Joslyn2001]. Here, [Ashby1958] described for systems with control that pursues a certain state with his law of requisite variety and his idea of variety destroying variety the relation between the variety of the state and the variety of influences. In his model, we have the system state for which control pursues to keep it as close as possible to a goal state. There are perturbations that influence the system state to deviate from the goal state. The controller has certain actions to counter the effect of perturbations. According to his law, a controller needs enough variety in their actions to counter the variety due to perturbations to the system in order to maintain the desired state. Equation 2.1 represents the relation between the system's influences and the goal variable (cf. [Heylighen and Joslyn2001]). V(S) represents the variety in the system state. V(P) represents the variety of perturbances. The variety of regulating actions is represented by V(A). B is a spontaneous decrease in variety due to buffering.

$$V(S) \geq V(P) - V(A) - B \tag{2.1}$$

Translating it to darts, we have - for example - the goal to hit the bull's eye. Hitting the bull's eye, the relevant part of the system state is the desired state. For hits close to the bull's eye the system state is closer to the desired state than for hits in distance to the bull's eye. Ideally, we have no or a very small variety in the system state always hitting the bull's eye field. Perturbations in dart can be inside the player, e.g., due to breathing or due to the heartbeat. They can also be in the environment, e.g., a sudden wind that influences the path of the flying dart. One action for the participant as a controller to counter perturbations is, e.g., to wait for the right moment to avoid perturbations due to heartbeat or breath. A buffer in this example is the bull's eye itself. Though it is small it covers a certain area allowing a variety of different positions or states to suffice the goal or to be the goal state. If we imagine that we increase the size of the bull's eye to the size of the dartboard then we see that for a fixed set of hits more hits are inside the bull's eye achieving the goal state and improving performance.

In our research questions, we ask for the differences between good and bad participants in darts, and the differences between participants that improve more than others. Now, we combine these with Ashby's law to concretize our hypotheses. Following Ashby's "variety destroys variety", we expect for good participants with lower variety in fulfilling the goal to observe a higher variety in throwing the dart in comparison with participants with a bad performance. For the development over time, we expect that participants that improve more in performance also show a bigger increase in variety during their throws compared to participants that improve less. So our approach is to compare variety in the regulation - represented by the variety in the physiology during the throw - with the variety in the fulfillment of the goal variable which relates to the inverse of performance. This leads us to the task of quantifying the physiology of throws as well as performance.

2.3 Approaches to quantify Physiology and Performance

We can quantify the throw process at different levels. We could measure the electrical activity for the brain to quantify cognitive processes. For the subsystem of the muscles that contract to generate the forces that lead to a movement, we could measure their electrical activity using EMG (electromyogram). The next level could be the kinematics of the throwing arm and hand - the forces and resulting accelerations, velocities and trajectories. Lastly, we could also track the kinematics of the thrown dart. Let's take a look at what others did in this area. [Cheng *et al.*2015] opposed a group of novice dart players with a group of expert dart players to find differences in electrical surface activity of the brain. They quantified physiology by measuring the electrical surface activity on the brain using EEG (electroencephalography) and extracting oscillations in a certain frequency interval. [Weber and Doppelmayr2016] explored the effects of imagining throwing on performance and electrical surface activity on the skull using EEG. Works that examine the kinematics of the throwing arm include [Lee *et al.*2014]. They explore the differences in throw movement due to alterations in the bone structure of the wrist after an injury. To collect kinematic data, they simulated the throw process stepwise taking pictures using CT (computer tomography) and reconstructed the movement from the collected data. Even though in our related work nobody used kinematics to quantify the physiology of a throw, we see some options to do so. [Preim *et al.*2009] do not investigate the movements of any kind of throw. But, they do investigate the dynamics of a fluid in the blood circulation. In their work, they give an overview of techniques to visually explore perfusion properties of body tissue. These properties are determined from tracking a fluid contrast-agent in the blood circulation using, e.g., MRI (magnetic resonance imaging) or CT. For each voxel of these recordings, they get a time-series on the measured signal intensity that changes due to the moving contrast agent. These series share a certain pattern when the contrast-agent crosses. In the analyses of dart throw kinematics, we have also process patterns that occur in every throw, e.g., the initial acceleration and the final deceleration of a throw. [Preim *et al.*2009] fitted a simple process model for the signal intensity due to the passage of the contrast agent in order to abstract the data and to extract certain form properties for the process in each voxel. Afterward, they aggregated these properties visually to show the distribution in the tissue. Having such properties for dart throws we may aggregate them to quantify variance across several throws.

Looking at performance measures, each of the following related works set the movement goal to hit the bull's eye by throwing the dart. [Cheng *et al.*2015] and [Weber and Doppelmayr2016] quantified performance as averaged ring scores over a batch of throws. They assigned decreasing scores from the bull's eye to the outer rings of the dartboard. Each hit was then awarded the score of the hit field. Closely related to ring scores are displacements of hits in relation to the bull's eye. While ring scores assign high values to hits close to the bull's eye and small values to hits in distance, it is the opposite for displacements. They measure the distance between the bull's eye and hit. So they are small when a hit was close to the bull's eye and big when far away. [Seidler *et al.*2017] investigated the effects of electrical surface stimulation on the brain to the adaption to changed visual conditions. They quantified performance as the mean of the horizontal displacement in cm. [van Beers *et al.*2013] used a different approach while investigating the learning rate in a darts experiment. They quantify performance using the variance in the displacement of hits and motivate this with the fact that the variance of hits must be small to reliably hit the bull's eye.

In the related work, each time just a single performance measure was used. We recognize here the problem of using just a single performance measure because for each one there are cases that seem equal according to the measure but that are different in fulfilling the goal on hitting the bull's eye. Regarding the use of variance as only performance measures, though a small variance in hits is needed to reliably hit the bull's eye, it is easy to see that this condition is not sufficient. Let's imagine a participant with the smallest possible variance in hits always hitting the same position. If this participant always hits the bull's eye then this shows the highest possible performance. Yet, when the hits are always at the lower rim of the dartboard then it fails the movement goal of hitting the bull's eye for which we expect a low performance. Nevertheless, the variance for both cases is equal. On the other hand, also regarding the displacement measures, we want to show an example pair of equal performance according to the displacement measure with different degrees in fulfilling the goal of hitting the bull's eye. For the first participant, we imagine a throw behavior of hitting repeatedly just the same ring on the dart board. Based on the mean displacement of this participant, we imagine a second one that shares the same mean displacement but whose hits are scattered over the dartboard.

We can now argue that the first one fulfills the goal less than the second because the first one is unable to hit the bull's eye while for the second one some hits may land inside.

We see that with just one of those measures we cannot completely quantify performance on hitting the bull's eye. In the following thesis, we do not want to search for a better measure for quantifying performance that overcomes this problem. Yet, we want to use for our analyses multiple complementary measures.

2.4 Approach to handle multiple Performance Measures

In the discipline of *multi-objective optimization* optimization problems with multiple performance measures are handled. **Objective** is the term for a performance measure in this discipline. Having multiple objectives, we may evaluate two elements for each of them. If all of those objective support that an element x is better than an element y then it is simple to conclude that x is better according to all objectives. However, if one objective supports that x is better than y and another objective supports the opposite - that x is worse than y - then we have a conflicting situation and no clear concept of better and worse. To be still able to apply the concept of better and worse, [Deb2011] propose to extract for a set of elements the non-dominated front which is a subset that we can consider better than the rest of the set. Extracting non-dominated fronts iteratively gives us a partition of the elements with multiple performance measures into disjunct subsets for which it defines that certain subsets are better than others.

The subsets are determined by iteratively computing the subset of non-dominated elements - also called non-dominated front - and removing it from the set for the next iteration. The non-dominated front is the set of elements for which we cannot state that they are clearly worse than any other element. Having multiple measures, we can define a domination relation for each pair of two elements based on the domination relations for single measures. For a single measure, an element x dominates the other one y when we consider x better than y (cf. Equation 2.2). If we think of scores, we might consider higher scores better than smaller ones. If we think of error measures, we usually consider smaller values to be better. For multiple measures, an element x dominates an element y when x is not dominated in any measure by y and x dominates y in at least one measure (cf. Equation 2.3). This relation allows that an element x either dominates an element y, that x is dominated by y, or that both do not dominate each other. Searching the elements that are not dominated by any other element gives us the non-dominated front (cf. Equation 2.4). The elements of the non-dominated front do not dominate each other but they dominate all the remaining elements. For this reason, we consider the elements of the this front to be better. By iteratively determining the non-dominated front and removing it from the set, we partition the original set into a series of fronts. For non-dominated fronts, all the elements inside a front do not dominate each other. Though, all elements of a front are dominated by elements of fronts that appear earlier in this series - except for the first front of the original set. For this reason, we can consider fronts that appear earlier in this series of fronts to be better than fronts that appear later. This gives us an order based on multiple performance measures.

$$
\begin{aligned}
&\textit{x dominates y regarding f} \iff \\
&\qquad x \underset{f}{\prec} y \iff \textit{x is better than y regarding f} \\
&\qquad\qquad \textit{with f as objective}
\end{aligned}
\tag{2.2}
$$

$$
\begin{aligned}
\textit{x dominates y} &\iff \\
x \prec y &\iff \textit{(x dominates y regarding at least one objective)} \\
&\qquad \textit{and (x is not dominated by y regarding any objective)} \\
&\iff (\underset{f \in F}{\exists} : x \underset{f}{\prec} y) \wedge (\neg \underset{f \in F}{\exists} : y \underset{f}{\prec} x) \\
&\qquad \textit{with F as set of objectives}
\end{aligned}
\tag{2.3}
$$

$$
\begin{aligned}
\textit{non-dominated front} =& \textit{subset of X consiting of elements that} \\
& \textit{are not dominated by any other element of X} \\
=& \{x | (x \in X) \wedge (\neg \underset{y \in X}{\exists} : y \prec x)\}
\end{aligned}
\tag{2.4}
$$

In order to visualize fronts based on 2 measures, we can exploit their inherent order. Since the elements of a front do not dominate each other, we know that for every two elements x and y the element x dominates y in one measure and y dominates x in another measure. Since there are just two measures, we can sort all elements in improving order according to one measure to a series that is sorted at the same time in degrading order according to the second measures. In Figure 5.3, Subfigure (2) we depicted two performance measures along the horizontal as well as the vertical axis. For each front, we used their inherent order to create a polyline which connects all elements. In the subfigure, the best fronts are located near the bottom and left part of the area. The worst fronts are in the top right.

Next, we take a look on how others applied non-dominated sorting for analyzing data using multiple performance measures. [Vrugt *et al.*2007] developed and analyzed a simulation on bird migration between Europe and Africa. They used flight time and energy use as performance criteria for different migration routes. They analyzed the relation between the first non-dominated front on these measures and the migration routes. They found that the tradeoff between both measures leads to two main branches in migration routes - each one prioritizing one measure. [Levin2014] explored the application of non-dominated sorting in the analysis of startup performance. They used growth and profitability as performance measures. By extracting the first non-dominated front, they were able to retrieve the relation between the measures that were used to generate their data. Both works focus on analyzing the relations in the first non-dominated front. Doing so, they ignore big parts of their datasets that are not optimal or even show bad performance.

[Hulikanthe Math2017] explored empirical data on startup performance using survival time and the number of employees as performance measures. With an analysis called intrafront-analysis, they opposed the ends of a given front from non-dominated sorting in order to identify differences between the ends or respectively tendencies along the front regarding other startup properties. They applied this to chosen good and bad fronts. In addition, they opposed a good and a bad front regarding their startup properties which they called interfront-analysis. [Hulikanthe Math2017] did not focus just on the first non-dominated front. Instead, using their intrafront- and interfront analysis, they also explored the worse parts of the data using multiple performance measures.

We see the possibility to use non-dominated sorting for our analysis using multiple indented performance measures on the performance in throwing darts. It will give us a concept of good and bad according to multiple measures. Based on this concept, we can oppose well and badly performing participants to find differences in their throw physiology. This will give us an answer to our research question that will be different from using just a single performance measure.

2.5 Goals and Expectations

In this thesis, our goal is to explore the relation between throw performance an throw physiology in darts. For this, we want to quantify the physiology of the throw movement. In addition, we want to quantify the performance of the throw movement. Here, we want to use multiple performance measures and analyze them in a combined way using the fronts from non-dominated sorting.

Due to the connection to Ashby's law of requisite variety, we expect better performing participants to show more variety in their physiology than worse performing participants. For the development over time, we expect participants that show a stronger improvement in performance to also show a stronger increase in variety in their physiology compared to players that improve less. Lastly, we expect to get additional findings by applying multiple performance measures and using them in a combined way for the analyses.

In the next chapter, we present our method that we developed to fulfill these goals and to test our expectations.

3 Data Basis

In this chapter, we describe the dataset that is the basis for our analyses. Based on our research questions and our expectations, our goal is to explore the relation between performance and physiology of dart throwing. We are grateful as we were provided with the dataset for a dart experiment which consists of data related to the outcome of throws and measures related to the physiology of the respective throws. First, we describe the experiment and the collected data. Second, we describe how the data is related to performance. Lastly, we describe why we focus in this thesis on the data for accelerating forces.

3.1 Dart Experiment

We thank Prof. Ross Cunnington, Tamara Powell, et al. from the University of Queensland and Queensland Brain Institute for providing us the datasets for the dart experiment that they conducted. For this experiment, they let participants throw darts subsequently 500 times with the task to hit the bull's eye.

They conducted this experiment at the University of Queensland and Queensland Brain Institute approved by the University of Queensland Medical Research Ethics Committee (UQ Project No. 2011001394). They conducted the experiment in two phases with **Phase 1** in 2015 and **Phase 2** in winter 2017 / 2018.

In a well-lit room they marked a line on the ground with a distance of 267cm to the wall with the dartboard marking the throw area for participants. In Figure 2.1, we see the dartboard that they used in this experiment. In the center, there is the red circular bull's eye. Around it, the dartboard is divided into 9 Rings and 12 sectors of a circle which creates 108 ring segment-shaped fields. The dartboard spans in total a diameter of 43cm. It was attached to the wall with a height between floor and bull's eye of 155cm in Phase 1 and 143 cm in Phase 2. As for darts, in Phase 1 two sets of 5-6 darts were provided. One set with a dart weight of 22g each as well as another with a dart weight of 26g each. In Phase 2 only a set of 5 darts with 22g was supplied.

They instructed each participant to stand behind the line marked on the ground, to try to hit the bull's eye, and to throw the set of 5-6 darts before retrieving the darts. In Phase 2, they gave participants additional information on how to throw according to the recommendations from `http://dartbrokers.com/dart-technique`.

In each recording session, there was just one participant. They conducted a session in one piece. In the beginning, they equipped participants with an inertial and EMG measurement device at the forearm as well as a mobile scalp EEG. The participants threw their set of 5-6 darts with own pace. In total participants threw 500 times separated into 5 sequences of 100 throws each. We call these sequences of 100 throws **blocks**. So block 1 refers to the first 100 throws while block 5 refers to the fifth or last 100 throws. After each block, there was a brief rest period for the participants.

During the experiment sessions, Cunnington et al. recorded the participant's physiological properties using the equipped sensors as well as they observed and recorded the hits on or off the dartboard.

They recorded for each throw the approximate position of the hit by recording the hit dartboard field using two measures **score** and **clock time**. The dartboard assigns to a hit in the bull's eye a score of 10. For each of the subsequent 9 rings, it assigns a score reduced by one. So it rewards the outermost ring with a score of 1. For missing the dartboard they assigned a score of 0. For clock time, we have 12 evenly spaced sectors going clockwise - in reference to a clock - having sector 12 starting on the top going clockwise until sector 1 starts. For each sector following clockwise the sector number increases by one. Hits in the bull's eye field do not have a clock time assigned. This clock time measure relates to the direction of the hit relative to the bull's eye. Except for the first 7 participants of Phase 2, they also recorded the clock time when participants missed the dartboard.

Previous to the throw session, they made the participants take part in a questionnaire to asses their familiarity with dart throwing asking them to estimate their personal frequency of dart games per year as well as the time that has passed since the last game. Immediately before the throw session, they

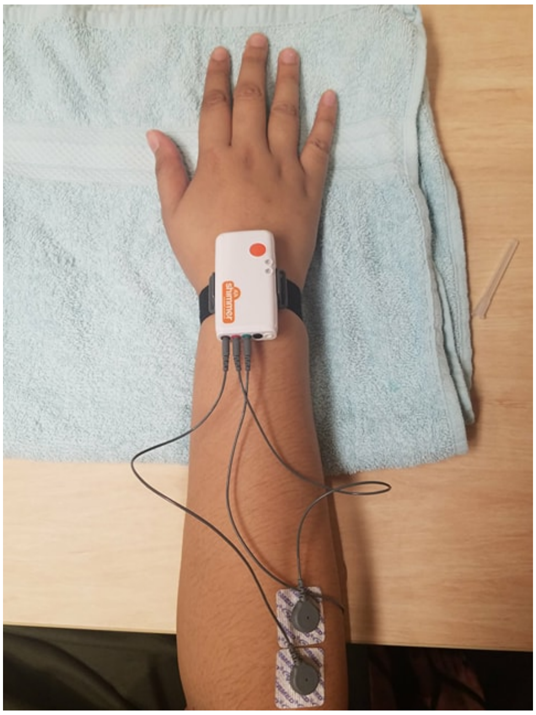

Figure 3.1: The inertial measurement unit attached to the wrist - Cunnington et al. attached the Shimmer 3 device by Realtime technologies ltd to the wrist of the trowing arm to record throw kinematics and muscle activity.

equipped participants with the inertial measurement unit (IMU) *Shimmer 3* by *Shimmer* at their right wrist for measuring characteristics of the throw movement (cf. Figure 3.1). The wrist is close to the fingers which hold the dart. For this reason, we expect that the measured influences are similar to the actual influences that affect the darts during the throw process. The IMU recorded during the experiment time series with a sampling rate of 512Hz for accelerating forces according to three axes relative to this device using an accelerometer, as well as rotation velocity in three axes relative to the device using a gyroscope and orientation using a magnetometer. As the measurements took place inside a building, we considered the magnetometer measurements not reliable from the beginning. The accelerometer settings were modified from a today unknown setting in Phase 1 to a setting of +/- 16g in Phase 2. In Addition, this device was extended by two bipolar electrodes for measuring electrical surface activity that we affixed on the arm above the extensor digitorum muscle.

This muscle takes part in releasing the darts. So, Cunnington et al. expected high electrical activity to be related to the moment of releasing the dart during the throw. As paid volunteers and having given written informed consent, 23 participants took part in Phase 1 as well as 23 other participants in Phase 2. There were 26 females and 20 males between 18 and 40 years with a mean age of 23 years. They had normal or corrected vision and no neurological or psychiatric disorders in the present nor in the past. All of them were right-handed and most of them were beginners in the game of darts.

As the first step of postprocessing, Cunnington et al. identified markers for individual throws in the recorded session data for further processing. Unfortunately, the EMG data was too noisy to identify single throws. In gyroscope data, they observed peaks in rotational velocity which correlate with throws. During a throw, the throwing arm moves faster than in the rest phases which also includes faster rotations. So, they scanned through the gyroscope data and identified those peaks and recorded for every throw the corresponding **peak time** as a marker that lays inside the timeframe of a throw. Furthermore, they discarded the datasets for participants with artifacts in EEG data or failed synchronization between EEG and IMU.

After all, we were provided with complete datasets consisting of hit data, kinematic data, and data regarding the electrical activity of the brain for 17 participants in Phase 1 and for 15 participants in Phase 2. Based on this data, we will quantify performance as well as the physiology of throws next.

3.2 Towards Performance Measures

In the experiment, for every throw, the approximate hit position was recorded using the measures score and clock time that describe the hit field. We take an initial look at the collected data. Figure 3.2 is a heatmap that shows the distribution of the hits on the fields of the dartboard and its surrounding sectors for the 100 throws in the first block of participant J14. The top and bottom of the diagram represent the top and bottom of the dartboard. Respectively, the left and right represent the left and right of the dartboard. We see that there are hits scattered all over the dartboard. Yet, the focus of hits lays in the lower half of the diagram mostly missing the bottom of the dartboard to the 4 lower sectors. Most hits are far from the bull's eye thus showing a bad performance.

In Figure 3.3, we see for a different participant that the distribution of hits is less scattered and less displaced from the bull's eye. This participant seems to perform better. The distribution for the next participant in Figure 3.4 is also more concentrated than the distribution for the first participant. On the other hand, the focus of the distribution is not centered around the bull's eye. It is displaced downwards. So, this distribution shows a performance worse than the second participant. Lastly, in Figure 3.5, we see a distribution that is also scattered all over the dartboard but without a strong tendency for a general displacement in a certain direction. Here, the performance also is worse than the performance of the second participant. To sum it up, we reach a higher performance when our hits are generally closer to the bull's eye and when our hits are less scattered.

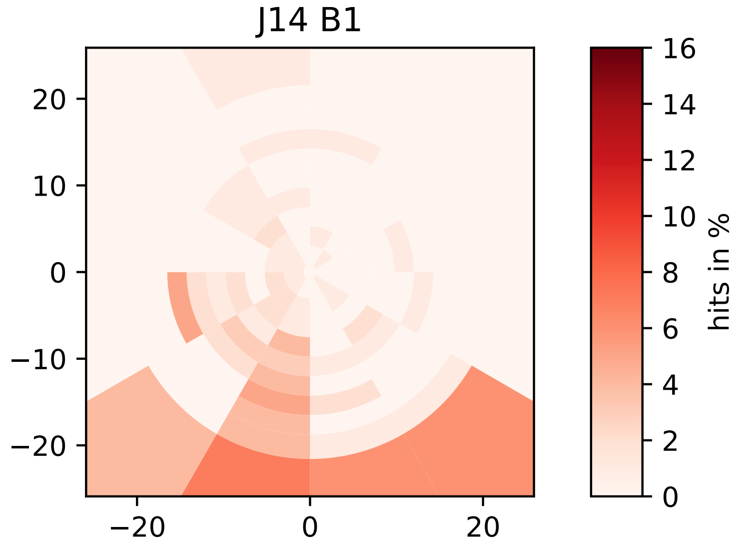

Figure 3.2: Distribution of hits with high scatter and high displacement - We depicted the distribution of hits on the fields of the dartboard and the surrounding sectors as a heatmap. This heatmap shows distribution of the 100 hits from the first block of participant J14.

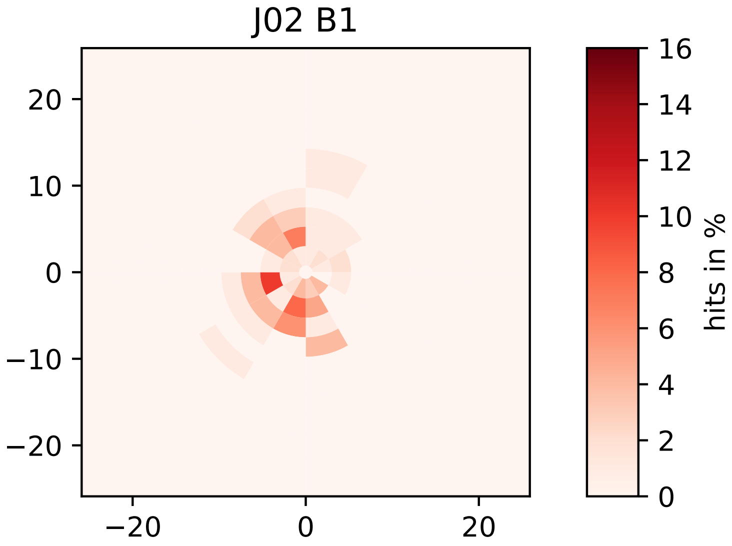

Figure 3.3: Distribution of hits with low scatter and low displacement

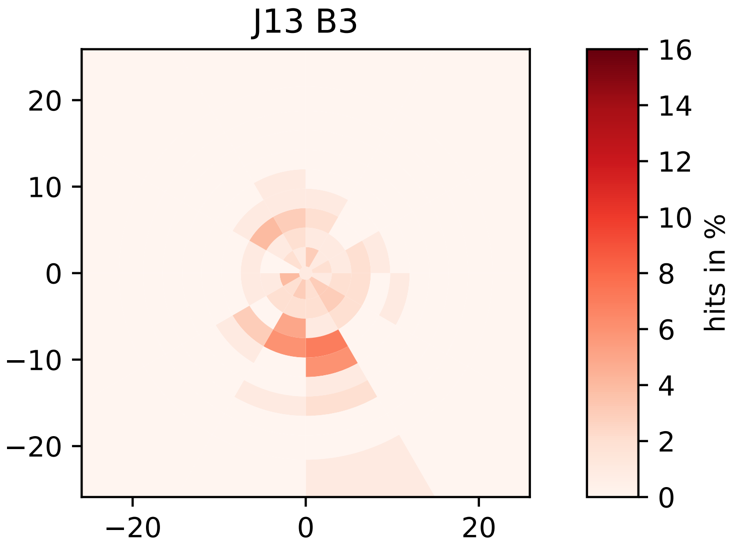

Figure 3.4: Distribution of hits with low scatter and high displacement

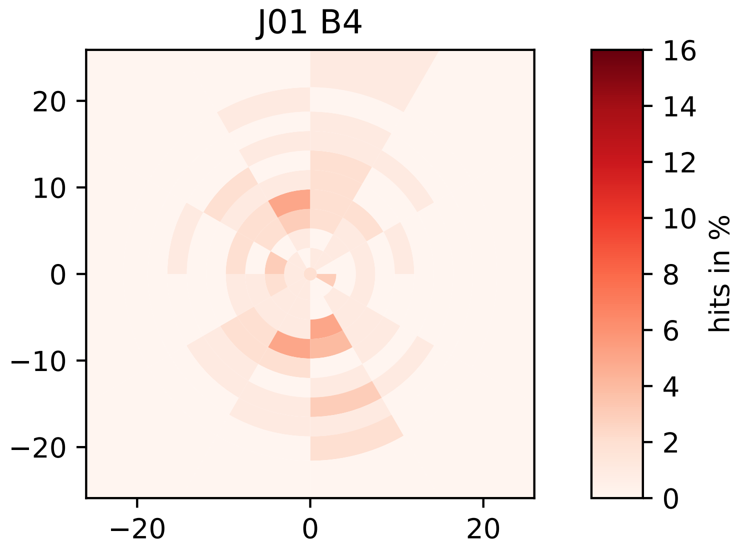

Figure 3.5: Distribution of hits with high scatter and low displacement

3.3 Focus on accelerating Force as recorded physiological Measure

In the experiment, Cunnington et al. recorded the way of throwing darts regarding the executed movements and the activity of the brain. For the throw execution, they recorded time series about the kinematic influences close to the thrown dart measuring the accelerating force and rotational velocity. Regarding the activity of the brain, they recorded time series on the activity of different brain areas using EEG. In the conceptual chain from brain activity, over muscle activity, over emerging movements, over flight trajectories, towards the fulfillment of the goal, we see more steps from brain activity to performance than from emerging movements towards performance. To begin analyzing the provided dataset, we decided to analyze the shorter connection from emerging movements to the fulfillment of the movement goal. For this reason, we ignore the EEG data in the following parts of the thesis. Instead, we focus on the kinematic data that describes the movements which emerged. For the kinematic data, we have the 3-dimensional time-series from the gyroscope, the accelerometer, and the magnetometer of the inertial measurement unit. We already dropped the measurements of the magnetometer for being not reliable due to being recorded indoors. For the gyroscope - which measures the rotational velocity relative to the device - and the accelerometer - which measures the accelerating forces also relative to the device -, we decided to focus further on the measurements of the accelerometer. Our reason is, that we see the measurements for accelerating forces to be more relevant towards the throw of darts because the rotational velocity depends on the technique used for throwing. In the case of a rotational throw - when the throw is executed mainly by a rotation in the elbow joint -, the rotational velocity correlates to a high degree to the velocity of the sensor and respectively to the velocity of the thrown dart. Yet, using a different technique like an ejecting throw - when the rotation takes place in the elbow joint and the shoulder joint -, weakens this correlation. Since we do not know which techniques the participants used, we decided to focus on the acceleration forces that affect the sensor instead. The accelerating forces of the sensor are similar to the accelerating forces that affect the dart while it is held. In Figure 3.6, we show the time series of two

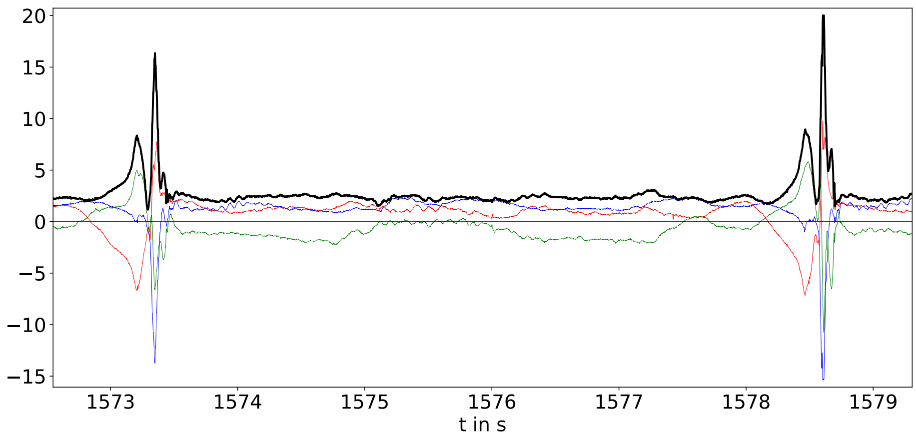

Figure 3.6: Time series for the accelerating force of two subsequent throws - We depicted the time series for the recorded strength of the accelerating force according of the three axes of the sensor in red, green, and blue. In addition, we added the time series for the norm of the accelerating force in black.

subsequent throws for the 3 axes of the sensor in red, green, and blue. First, we see for both throws after the time 1573 and after the time 1578 distinct amplitudes. Second, also between both throws, the amplitudes do not collectively get close to 0 though it is a phase of slow motions and small accelerations compared to throws - so-called a **rest phase**. Between throws, participants usually rest their arm and hand or take the next dart. A second accelerating force that the accelerometer recorded compound with accelerating forces to accelerate the dart is gravity. Gravity is a static accelerating force that is directed downwards. For each time of the recorded series of accelerating forces, we have three components according to the three measuring axes of the sensor. Together, these components represent a force vector relative to the sensor. We can compute the length or strength of this force vector using the Euclidean **norm** (cf. Equation 3.1).

$$
\begin{aligned}
\|\vec{x}\| = & \sqrt{{x_1}^2 + {x_2}^2 + {x_3}^2} \\
& \textit{with } \|\| \textit{ as Euclidean norm} \\
& \textit{with } \vec{x} \textit{ as vector} \\
& \textit{with } x_1,\ x_2, \textit{ and } x_3 \textit{ as vector components}
\end{aligned}
\tag{3.1}
$$

We aggregated using the norm the three channels for every measurement time to create a time series for the norm. We depict this series as the black series in Figure 3.6. The black series shows us for the rest phase between both throws a certain baseline that is shifted to the value of 2.5. The almost constant strength of the accelerating forces in the rest phase matches to the constant strength of gravity with only small accelerations due to movements. Though the strength of the accelerating forces is almost constant in the rest phase, we also see that the amplitudes for the separate channels change between both throws. For example, the value for the green series raises, drops, and raises again almost in parallel to the blue series. The reason for this is our sensor and its measuring axes that rotate during movements while the gravity remains pointing downward. This also affects the accelerating forces that we are interested in to quantify the physiology of the throw movement. So due to this rotation, a constant accelerating force may initially appear in one channel, may be spread over time to several channels, and may, in the end, remain in a different channel. The norm is independent of the rotations of the force vector because rotations do not change the length of the vector. For this reason, we decided to focus further on the norm of the accelerating forces to quantify the physiological properties of throws.

Yet, another observation that we made is that the scales for the accelerating forces for the records of Phase 1 and Phase 2 data do not match. In rest phases, the accelerating forces due to movement should be small while the main accelerating force is gravity with a constant strength. For this reason, we expect the average value and the scatter of values in rest phases to be similar. In the time series of the norm of accelerating forces for all five blocks of throws for 9 participants from Phase 1 and 9 participants from Phase 2, we selected manually each time three rest phases. For these 54 rest phases, we determined the mean and standard deviation to characterize the position and the scatter of the values. In Figure 3.7, we confront the mean along the horizontal axis and the standard deviation along the vertical axis. Every point represents the properties of a single rest phase. The blue points for the rest phases of Phase 1 form a small cluster in the lower right with small values for mean and standard deviation. The orange points for the rest phases of Phase 2 are clearly separated from this cluster having higher mean values. We consider the rest phases to be corresponding reference values of the scales that were used to measure the accelerating forces. In the figure, we see that these references values do not correspond well between Phase 1 and Phase 2 which implies the application of different scales. Thus, we cannot compare the amplitudes between both phases because any value has different meanings according to the different scales, e.g., a participant with the height value 190 is big according to the centimeter scale but small according to the millimeter scale. As for this reason the amplitudes are not comparable and as the scales in Phase 2 vary more, we decided to drop the Phase 2 data and focus on the data for the 17 participants from Phase 1.

In consequence of our decisions, we focus on the norm of the accelerating forces for the participants in Phase 1 in order to quantify the physiology of throws.

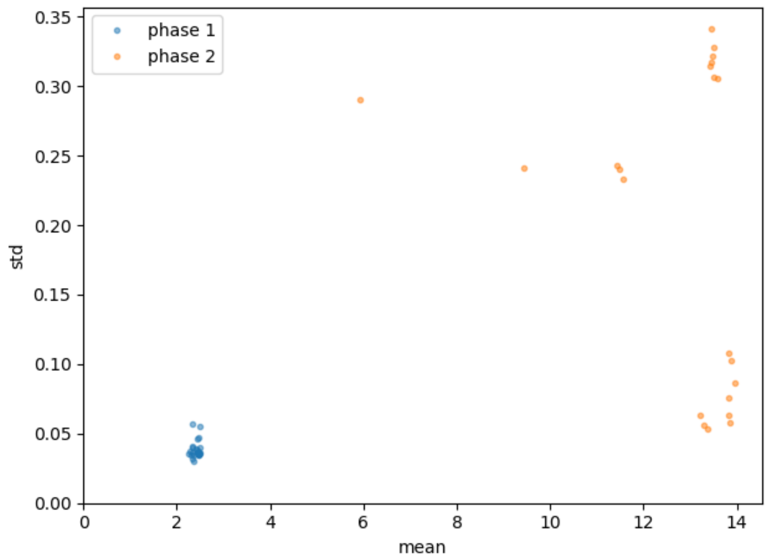

Figure 3.7: Properties of rest phases - We manually selected 3 rest phases in the time series for the norm of accelerating forces for each of 9 participants from Phase 1 and 9 participants from Phase 2. For the selected rest phases, we determined the mean and the variance. We confront these properties with the mean along the horizontal axis and the standard deviation (std) along the vertical axis.

4 Proposed Method

In this chapter, we want to describe our developed method.

Based on our research questions and our expectations, we want to explore the relation between performance in dart and the way of throwing the dart which we call the physiology of throws. The main goal of our method is to provide information to explore the relation between performance and throw physiology. To do so, we quantify performance and throw physiology, as well as the development over time in those measures. Finally, we analyze in Sections 5.2 and 5.3 the developed visualizations that oppose both sides of throwing dart, which we used to characterize the relation.

4.1 Performance Measures

In the provided experiment, the outcome of every throw was recorded using score and clock time that represent the approximate position for the hit on the dartboard. Now, we want to use these measures to quantify the throw performance of participants. In Section 2.3, we decided to use multiple performance measures to quantify the performance of throwing dart which is different to the corresponding related work. So, we observed in Section 3.2 that we consider a distribution of hits better regarding the goal of hitting the bull's eye when the distribution is less scattered and generally shows a smaller deviation from the bull's eye. In order to capture both - the scatter of hits and a general displacement -, we define two performance measures. We define both measures as error measures, i.e, the smaller the error the better the performance. So, an error of 0 represents the maximum possible performance according to the respective measure. First, we name the concept of performance from the general proximity of hits towards the bull's eye **accuracy**. The participant in Figure 3.2 showed a smaller accuracy than the participant in Figure 3.5 because the hits of the first one generally deviated from the bull's eye downwards while the hits of the second one are generally centered at the bull's eye. Based on accuracy, we define the **accuracy-error** as the distance between the center of the bull's eye and the center of the hits (cf. Equation 4.1). As we will see in an instant, we can remove the position of the center of the bull's eye from the equation.

$$\begin{aligned} \textit{accuracy-error} =& \|\vec{x}_{\textit{center of bull's eye}} - \vec{x}_{\textit{center of hits}}\| \\ =& \|\vec{0} - \frac{1}{|H|}\sum_{\vec{x}\in H}\vec{x}\| \\ =& \frac{1}{|H|}\|\sum_{\vec{x}\in H}\vec{x}\| \\ & \textit{with } \vec{x} \textit{ as position vector for a hit} \\ & \textit{with } H \textit{ as sequence of hits} \end{aligned} \tag{4.1}$$

For the second performance measure, we name the concept of performance from smaller scatter of hits **precision**. The participant in Figure 3.3 has a higher precision than the participant in Figure 3.5 because the first one shows a distribution of hits that is more concentrated than the distribution of the second one. Based on precision, we define **precision-error** as the mean distance between two hits (cf. Equation 4.2).

$$\begin{aligned} \textit{precision-error} =& \frac{1}{|P|}\sum_{\{\vec{x}_1,\vec{x}_2\}\in P}\|\vec{x}_1 - \vec{x}_2\| \\ & \textit{with } \vec{x}_1 \textit{ and } \vec{x}_2 \textit{ as position vectors for hits} \\ & \textit{with } P \textit{ as sequence of all pairs of hits} \end{aligned} \tag{4.2}$$

For the definition of accuracy-error and precision-error, we assumed that we already have the exact positions of the hits. Yet, in our experiment, we collected just an approximate position consisting of score and clock time that describe the hit field. Thus, we estimate exact hit positions from the dimensions of the dartboard and its fields relative to the center of the bull's eye with the following strategy. For hits in the bull's eye field, we mapped the position to the center of the bull's eye with coordinates $\binom{0}{0}$. For hits in ring segment-shaped fields, we mapped the position to the intersection of the inner arc and the angle bisector of the respective ring segment. For misses, we mapped the position to the intersection of the outer rim of the dartboard and the angle bisector of the respective sector. Since all estimated positions are relative to the center of the bull's eye at $\binom{0}{0} = \vec{0}$, we can remove the center of the bull's eye from Equation 4.1 for the computation of accuracy-error. In addition, as we use the original dimensions in cm, the values for both performance measures also have centimeters as units.

In Table 4.1, we show the values for accuracy-error and precision-error for the examples from Section 3.2. As before, the second distribution is better than the first distribution according to both measures representing less scatter and general less deviation from the bull's eye. The third distribution is better than the first according to precision-error being less scattered but worse than the second distribution according to accuracy-error having generally a bigger displacement from the bull's eye. The fourth distribution of hits is worse than the second and the third according to precision-error being more scattered. Yet, it is better than the first distribution according to accuracy-error. The fourth distribution shows the smallest accuracy-error being better than both distributions with low precision-error.

Table 4.1: Performace measure examples - We computed for the examples from Section 3.2 the values for accuracy-error and precision-error rounded to one position after decimal point.

	figure	accuracy-error	precision-error
1st	3.2	8.5	14.4
2nd	3.3	1.2	5.1
3rd	3.4	2.3	7.2
4th	3.5	0.6	11.2

In order to quantify the performance for a block of throws, we estimate the exact positions of hits which we use to compute our performance measures accuracy-error and precision-error.

4.2 Describing the Throw Pattern by Throw Features

In Section 3.3, we decided to focus on the norm of accelerating forces as a measure to characterize the physiology of throws. In this section, we describe how we extract specified features from single throws using a model regarding the physiological processes and how we aggregate those features over many throws to quantify the physiology of throwing.

To prepare the data, we determine the norm of accelerating forces and extract the time series for every throw. For every throw, we use the peak time as a reference point in time and extract an interval starting 1500 milliseconds before the peak time and ending 500 milliseconds after it. In Figure 4.1, we see the extracted series for all throws of participant J01. We recognize a characteristic pattern consisting of to subsequent peaks for the processes during throwing. As we can see a clear pattern for the curves of the time series, we follow the

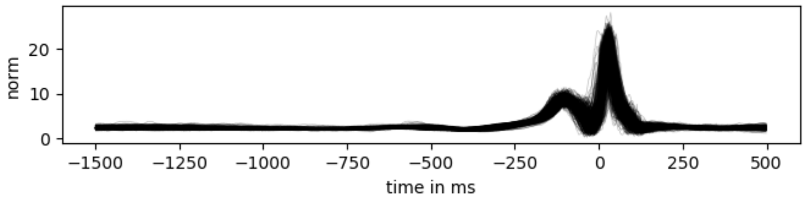

Figure 4.1: Extracted time series of the accelerating force for all throws of participant J01

idea of [Preim *et al.*2009] to simplify the time series to few form features that describe a single throw. To do so, we develop a model which we match to the curve of a throw to extract characteristic throw features.

These features base on a conceptual model of the throw process. In Section 2.1, we already described the process of throwing a dart consisting of a preparation phase to move the dart to an initial throw position, an acceleration phase to accelerate the dart to a certain velocity, and a deceleration phase to stop the throwing arm. The movements in the preparation phase are slow and for this accompanied by weak acceleration. In the acceleration phase, the velocity of the arm increases a lot in a short time. So here, the acceleration is strong. Lastly, in the deceleration phase, the velocity of the arm decreases in a short time a lot. So in this phase, there is a strong deceleration. Hence, the acceleration phase and the deceleration phase with their strong acceleration correspond to the first and the second peak that we see in Figure 4.1.

The model that we fit to the curve is piecewise linear, i.e., it consists of linear and constant functions that apply in a certain interval. We can see in Figure 4.2, that we approximate the baseline in the rest phases before and after throws using a horizontal line. Using a raising and a falling line segment, we model each peak for acceleration and deceleration phase. Using this model, we

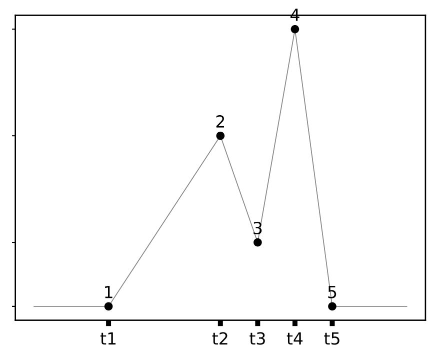

Figure 4.2: Time related form features in the piecewise linear model

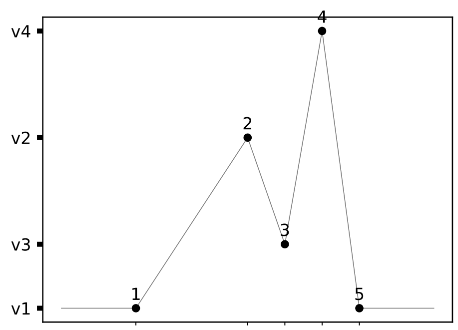

Figure 4.3: Amplitude related form features in the piecewise linear model

gather 5 points in time at the connection of the line segments to characterize the model which we show in Figure 4.2. We describe with **t1** the begin of the acceleration phase, with **t2** the time of the peak acceleration, with **t3** the time of the transition towards the deceleration phase, with **t4** the time of the peak deceleration, and lastly with **t5** the end of the deceleration phase. On the other hand, we have 4 amplitude related features at the connection of line segments. With **v1**, we describe the amplitude of the baseline, with **v2** the amplitude of the peak acceleration, with **v3** the amplitude at the transition towards the deceleration phase, and with **v4** the amplitude of the peak deceleration. There is no need for a v5 since at the end of the deceleration phase we reach the baseline whose amplitude is already described by v1. To limit the valid configurations of those form features, we defined a set of constraints including that the time-related form features keep their order (cf. Equation 4.3) and that the amplitude-related form features maintain both peaks (cf. Equation 4.4).

$$t1 < t2 < t3 < t4 < t5 \tag{4.3}$$

$$\begin{aligned} v1 &< v2 \\ v2 &> v3 \\ v3 &< v4 \\ v4 &> v1 \end{aligned} \tag{4.4}$$

In order to match our model as good as possible to the time series of a throw, we construct first a good initial solution for the model and optimize all form features afterward using the Levenberg–Marquardt algorithm (see [Ranganathan2004]) that depends on a good initial solution as it finds local optima. To quantify the distance between data and model, we define our residual as the sum of squared differences between data and model over the whole time series for the throw. In Figure 4.4, we show the time series for the norm of the accelerating forces in black. To find a good initial solution, we start with a static predefined model that we scale to match the maximum amplitude in the data. Afterward, we search with a brute force approach the best temporal offset. These three intermediate models are depicted in the figure with grey dotted lines. Finally, we apply the Levenberg-Marquardt algorithm to adjust all form features to minimize the residual value. The final model, we depict as red polyline in the figure. We see that it could capture the baseline as well as the peaks though there is some noise that our model cannot represent.

At this point, we have a set of form features that describe the shape of the time series for a single throw. Yet, the time-related features are still relative to the peak time which may vary relative to the process in the throw. The amplitude-related features still contain the influence of the gravity which is not part of the accelerations due to the throw movement. To avoid those influences, we define **throw features** which base on the differences between form features. From all combinations of the 5 time-related features, we define 10 differences as time-related throw features. With the notation d_t1_t2, we refer to the

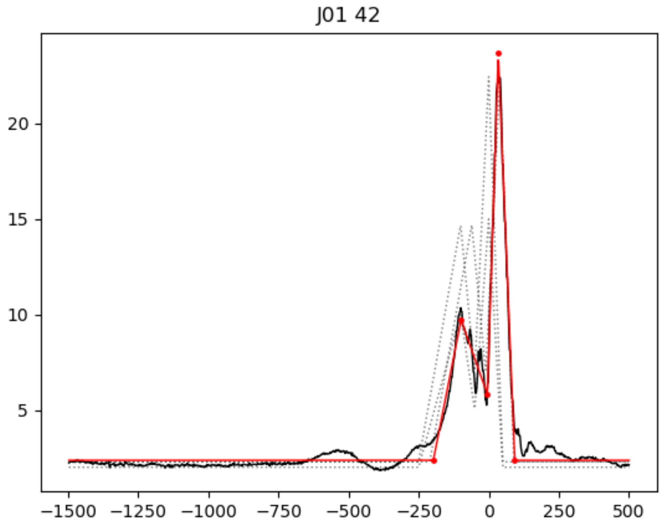

Figure 4.4: Fit for the 42nd throw of participant J01 - We depicted the time-series as the black line, the representation of the final fitted model as the red line, and the intermediate results for finding a good initial model for fitting as grey dotted lines.

difference of t1 to t2 - in other words, $d\_t1\_t2 = t2 - t1$. In Figure 4.5, we see a representation of d_t1_t2. So, it represents the duration from the begin of the acceleration phase to the time of the peak acceleration. From all combinations of the 4 amplitude-related form features, we define 6 differences as amplitude-related throw features. Respectively, with the notation d_v1_v4, we refer to the difference from v1 to v4, i.e., $d\_v1\_v4 = v4 - v1$. This throw feature represents the amplitude of the peak deceleration without the height of the baseline as we can see in Figure 4.6. Lastly, we define as 4 additional throw features the 4 slopes for the acceleration and deceleration phase. With the notation s12, we refer to the slope between the first and the second connection (cf. Figure 4.7), i.e., $s12 = d\_v1\_v2/d\_t1\_t2$.

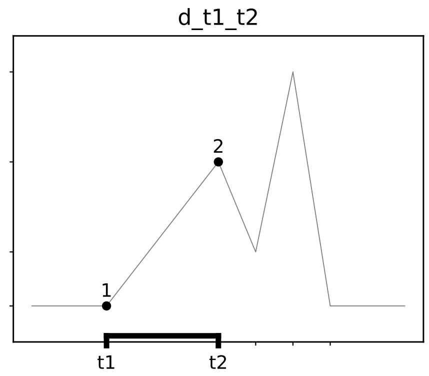

Figure 4.5: d_t1_t2 as example for a throw feature from a time difference

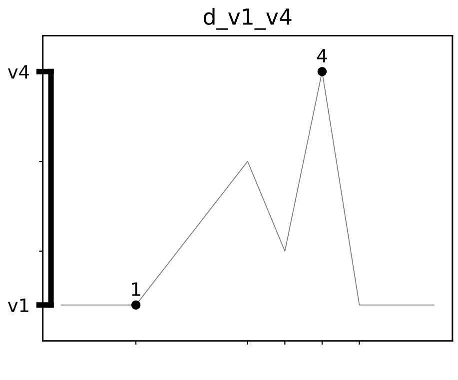

Figure 4.6: d_v1_v4 as example for a throw feature from a amplitude difference

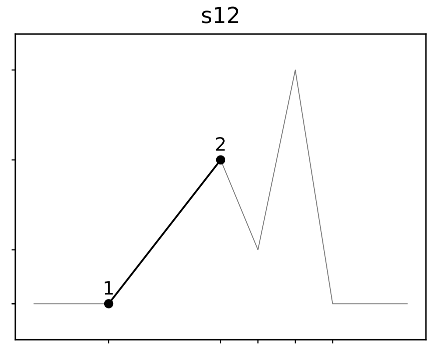

Figure 4.7: s12 as example for a throw feature from a slope

Having those 20 throw features, we can characterize throws individually. Yet, we have no measure for the variance of the movement in general. To characterize general properties of the throwing of a participant, we determine properties of the distribution for each throw feature over a block of throws. We characterize the scatter or variance of each feature using the **IQR** (interquartile range). For the average value of each feature, we determine the **median**. We chose both for the reason that they are more robust against outliers that may occur in the fitting.

In order to quantify the physiological properties of throwing, we abstract the time series data of accelerating forces by fitting a model from which we extract throw parameters that characterize each throw individually. Afterward, we determine for each throw feature the distribution properties median and IQR during a block of throws to describe the general value and the scatter.

4.3 Quantifying Development

We have now performance measures that describe in the time frame of a certain block the status of performance. In addition, we have distribution properties of throw features that describe also in the time frame of a certain block the way of throwing dart. As we are also interested in the changes over time, we quantify the development as follows. We compute the differences for performance measures as well as the distribution properties. We set the first block as reference value computing the differences in all these measures from the value regarding the first block to the value of later blocks, e.g., $development_{accuracy\_error} = block2_{accuracy\_error} - block1_{accuracy\_error}$. As we have 5 blocks in the experiment we will have for one measure 4 differences - to the second, third, fourth, and fifth block.

5 Results

In the following sections, we describe in detail the analyses that we make to answer the research questions as well as the corresponding observations. First, in Section 5.1, we have a closer look at the performance measures that occurred during our experiment and which are the basis for the confrontation in the following analyses. In Section 5.2, we look for differences in our throw features based on confronting good and bad groups based on the performance measures. In Section 5.3, we look for differences in the development of our throw features.

In order to explore the relation between performance measures and physiological measures of throws, we follow the idea as seen in the work of [Günnemann *et al.*2010]. They select groups of similar elements according to one space and visually explore their distribution in a different space with the help of scatter plots. Based on these visualizations, they identify manually patterns of interest. We select a good and a bad group according to the performance measures and explore their relation according to the distribution properties of throw features.

Regarding our first expectation - that participants with a better performance have a higher variance of throwing -, we conduct three analyses. We select both groups first based on accuracy-error (see Section 5.2.1) and second based on precision-error (see Section 5.2.2). Third, we select them according to the non-dominated fronts on both measures in the manner of an interfront analysis (see Section 5.2.3). We do not conduct an intrafront analyses because two groups based on the opposite ends of front have no notion of good and bad, or better and worse. Regarding our second expectation - that participants that improve more also show a stronger increase in the variance of throwing -, we conduct another three analyses. First and second, we confront groups based on the development in either accuracy-error (see Section 5.3.1) or precision-error (see Section 5.3.2). As third and last analyses, we select the groups based on the non-dominated fronts on the development in both performance measures as interfront analysis (see Section 5.3.3).

5.1 Performance measures from Experiment Data

In the following two subsections, we want to get an impression on the occurring values for performance measures that we receive from applying our performance measures to the experimental data. In addition, we want to get an idea of how the characteristics of these values will influence the following analyses for answering our research questions. In the first subsection, we will have a look at the actual values for accuracy-error and precision-error which we introduced in Section 4.1. In the second subsection, we will examine the development in those measures.

5.1.1 Regarding Block Performance

We aggregate the performance data from our experiment for single hits - score and clock time - for each of the five blocks for each participant computing the values for precision-error and accuracy-error. These represent respectively the mean distance between two hits on the dartboard in cm and the distance of the center of the hits to the dartboard center in cm. Since both measures are error measures and it is better to make smaller errors, we consider smaller values better than higher values. Or in other words, we consider low values - relating to the range of occurring values - as **good** and the ones with high values as **bad**. The resulting values for the performance measures are shown in the subfigures of Figure 5.1.

Subfigure (0) shows us a histogram regarding accuracy-error where we get an impression on the distribution of occurring values. We see the range starting from the left close to zero ending on the right close to 10. We see this distribution to be skewed towards the left or the better values with the majority of data points having small values for accuracy-error. Regarding precision-error, we see the corresponding histogram in Subfigure (1). The occurring values for precision-error range from 5 up to 20. In a different way from accuracy-error, this distribution is slightly skewed towards the right, i.e., towards higher

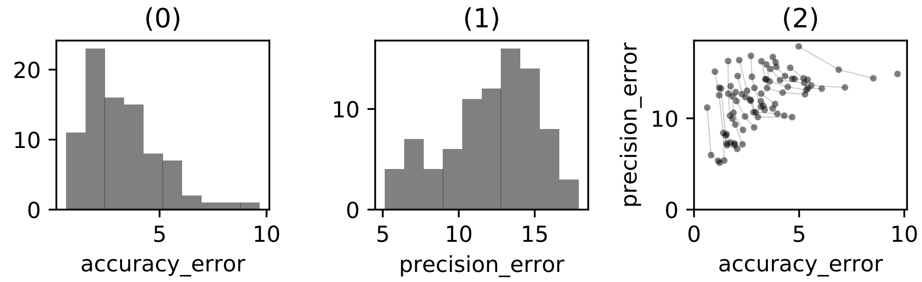

Figure 5.1: Occurring performance measure values - We determined our performance measures for each block of each participant using the data from our experiment. Subfigures (0) and (1) are histograms on this data showing the frequency of occurring values. Subfigure (2) is a scatterplot confronting both measures and also showing the non-dominated fronts as polylines.

values for precision error. Comparing Subfigures (0) and (1), we observe as an additional difference that for accuracy-error some participants in certain blocks were able to reach error values close to 0. Yet, for precision-error, we see nobody with a precision-error less than 5 in any block.

From these two subfigures, we get an idea for the distribution of one performance measure independent from the other one. The opposing skewness of both distributions may be a hint towards a negative correlation, i.e., that low values for accuracy-error may occur together with high values for precision-error.

In order to get an impression on the relation between both measures in our experiment, we look now at the remaining subfigure - Figure 5.1, Subfigure (2). This diagram is a scatterplot confronting accuracy-error along the horizontal axes with precision-error along the vertical axis. Each point still represents a certain block of a certain participant. In addition, we see the fronts from non-dominated sorting as polylines connecting the elements that belong to it. We explained how we determine these fronts in Section 2.4.

Looking at the data points, we see that they cover for accuracy-error a range starting from left close to zero going to the right up to a value of 10 like before in Subfigure (0). Also in precision-error, they cover the range starting at the bottom with about 5 going up to a value close to 20 just like in Subfigure (1). However, we see no negative correlation which would be visible as a perceivable line or strip of points going from the top left to the bottom right of the diagram. What we see resembles more a positive correlation. It is a **triangle**-like distribution of points starting at the bottom left with performances good in accuracy-error and precision-error going up to performances still with good accuracy-error but now with bad precision-error values and ending in the top right with performances that are bad according to both measures. Regarding the non-dominated fronts, the best are located on the left and the worst in the top right. The best fronts with low accuracy-error mainly go vertical from performances with high precision-error to performances with low precision-error whereas the worst fronts follow a direction that is diagonal to horizontal. Underneath the diagonal of this observed triangle are no data points. The area that would complete the upper triangle of data points to a rectangle, we refer to it in the following as **empty triangle**. This empty triangle means that in our experiment and according to our method no participant showed a performance in any block that would lead to a combination of low values for precision-error and high values for accuracy-error.

Observing this empty triangle, we wonder for the reason of this shape. One possible reason is the errors that we introduce in our method that may distort the distribution of occurring values. We measure the position of the hit of a dart on the dartboard using the dartboard field that the dart hit. Later, we estimate for each hit a position in these fields independent of the actual position in the respective field in order to apply our performance measures. Here, we introduce an error between the actual position of hit and the estimated position. Since the dartboard fields increase in area from the inside to the outside, participants may be affected to a different extent for example according to their accuracy-error. So participants with higher accuracy-error - a bigger offset of their hits from the center - are affected by bigger errors as they tend to hit the bigger fields more often than participants with low accuracy-error. This distortion may move occurring performances with low accuracy and high precision above the diagonal of the triangle where we cannot distinguish them anymore from the other performances. Since we determine the fronts based on these values they may not capture the initially intended combination of elements that are exceptionally good in just one performance measure and elements that are good in both measures. This would influence the further analyses based on these fronts.

In order to test whether our method is the reason for the empty triangle or in other words whether it is possible to achieve performances in it, we visualized the distortions according to both measures. Following the idea of what would be different if we had the exact offset in horizontal and vertical direction instead of clock time and score, we made a simulation to generate artificial performance data. For this, we generated artificial performance data to confront performance measures that contain the quantification to the fields of the dartboard and performance measures that do not. We chose simple circular two-dimensional normal distributions on the plane of the wall as representatives for the distribution of hits relative to the dartboard center from which we drew samples as hits with x and y relative to the dartboard center. For these distributions, we have 2 parameters - one for the location or center of the distribution and one for the extent of horizontal as well as vertical scatter - the standard deviation. Since the dartboard is rotational invariant, we consider for the location parameter just a positive offset to the right of the dartboard center. We defined 500 different distributions based on the combinations of 50 different values for the horizontal offset to the center and 10 different values for the standard deviation. From each of these distributions, we drew a sample block, i.e., 100 samples as simulated hit positions. For each of these artificial blocks, we computed the performance measures using the actual hit position from sampling for a first version and using the quantified positions from mapping to score and clock time for a second version. If the distortions that lead to the empty triangle are a general property of our performance measures we should observe these distortions and the empty triangle when we look at the first version of the performance measures. If these distortions are a result of our process - of using the dartboard and estimating the position of hits - we should observe the empty triangle just for the performance measures from the artificial data that was mapped to the dartboard fields.

Similar to Figure 5.1, Subfigure (2), we confront in Figure 5.2 accuracy-error along the horizontal axis and precision-error along the vertical axis. Here, we marked the triangle-like area that contains performances that occurred during our experiment with a continuous line. The empty triangle is marked with a dashed line. For each of the 500 distributions that we sampled there is an empty circle that represents the performance measures that are directly based on the sampled hit positions. For the empty circles with precision-error close to 0, we see the regular distances that we chose for the values of the offset parameter. The more we look up the harder it gets to recognize these lines. In the dashed triangle representing the empty triangle, we see lines of empty circles. So just according to our performance measures, it is

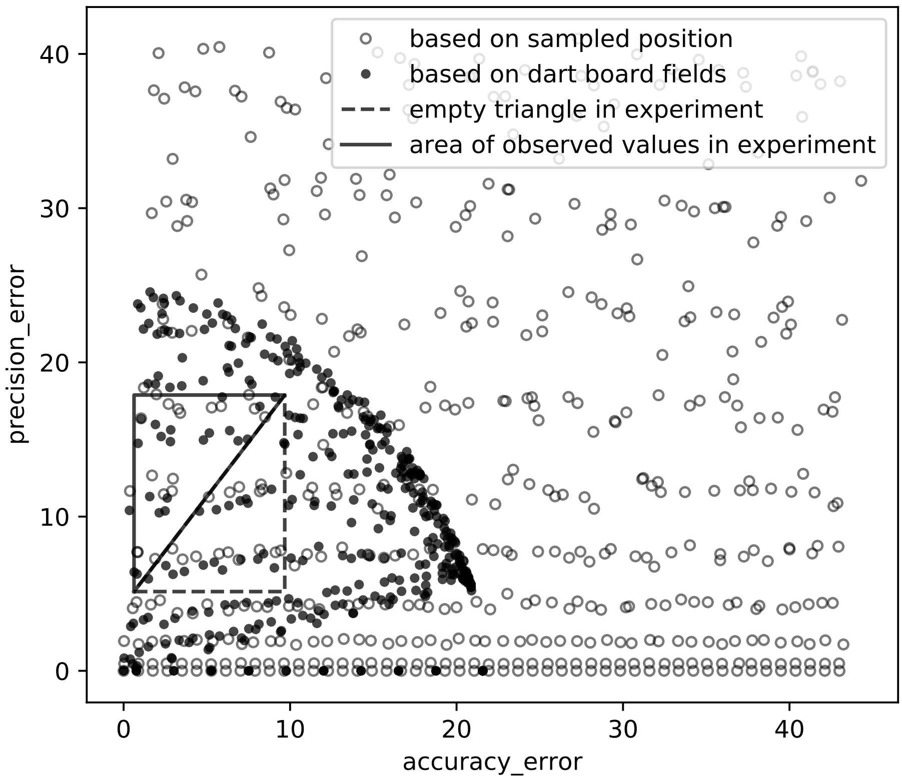

Figure 5.2: Visualizing distortions in the relation between our performance measures - We generated artificial blocks of hit positions that we used to compute the performance measures directly (empty circles) as well as by mapping them to the fields of the dartboard first (filled circles). We marked as a reference the area that contains the occurring values in our experiment as well the empty triangle underneath.

possible to reach this area. For each of the sampled distributions, there is a filled circle in the figure which represents the performance measures based on score and clock time. For these, we see multiple distortions. First, we see that the distribution is limited to the top and the right by a bow-like structure. Most likely this originates from the limited size of the dartboard and the mapping of missing hits to the margin of the dartboard. For example, if participants always miss the dartboard to the right then it makes no difference how far they miss as all these hits are later mapped to the right margin of the dartboard. According to this fact, we also see the maximum in accuracy-error for the filled circle at about 21 cm which is the radius of the dartboard. The second distortion is a positive correlation that we see for precision-error below 5. That fits the notion that the error that we introduce gets bigger to the outer rings. Nevertheless also in the area of the empty triangle, we see filled circles. Comparing corresponding filled and empty circles inside the triangle for the occurring values from our experiment, we see the filled circles shifted downwards. This means that using the dartboard for measurement, we underestimate precision-error.

In summary, we observed between our two performance measures from our experiment a triangle-shaped relation that is - according to our simulation - neither caused by their definition nor caused by introducing quantification errors by using the dartboard.

5.1.2 Regarding Performance Development

In order to quantify development with respect to the performance measures, we determine the differences in the performance measures from first blocks to corresponding later blocks. A negative difference for example in accuracy-error tells us that the accuracy-error in the later block was smaller than in the first block. This means that the corresponding participant **improved** in accuracy-error respective to his accuracy-error in the first block. Correspondingly, a positive difference refers to an increase which means that the participant **degraded**. Apart from improving and degrading, we can also observe that participants show no or hardly any development with difference values close to 0. Applied to the performance measures from our experiment, we determine for each participant the differences from the first block to the later 4 blocks.

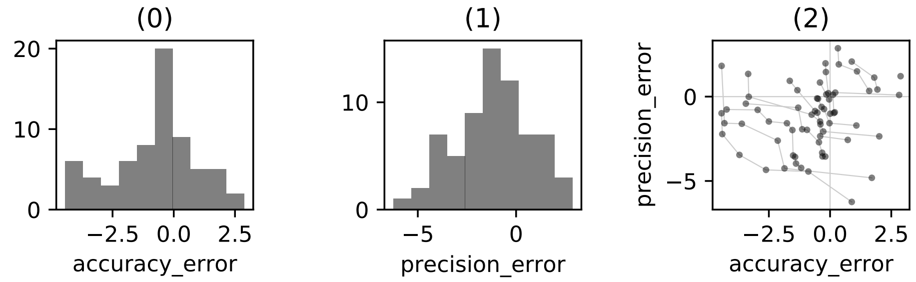

Figure 5.3: Occurring values for development in performance measures - We determined from the occurring performance measure values the differences from the first to the later blocks for each participant. Subfigures (0) and (1) are histograms on this data showing the frequency of occurring values. Subfigure (2) is a scatterplot confronting the differences in both measures and also showing the non-dominated fronts as polylines.

We visualize these differences in the subfigures of Figure 5.3. Subfigure (0) is a histogram that shows us the distribution of occurring difference values regarding accuracy-error. The distribution ranges from the smallest values - improvements of around -5 - on the left to the biggest values - degradations up to 2.5 - to the right. We observe the highest peak of the distribution to cover negative values but also to show very small development close to 0. Also, looking at the area left and right of 0, we see that the majority of occurring difference values are negative. For accuracy-error, we see that participants improved more often and also to a bigger extent than they degraded. So, we see a tendency of getting better in accuracy. Switching to precision-error, Subfigure (1) shows us accordingly a histogram for the distribution of occurring difference values precision-error. Similarly, the values range from -5 up to 2.5, the peak covers negative values close to zero, and occurrences have predominantly negative values. Yet, this time, the peak is not as pronounced as before. Also for precision-error, we see that participants improved more often and to a bigger extent which we interpret as a tendency of improving in precision-error over time. Even though both distributions are similar, yet, we do not know the relation between the development in both measures. For this, we look at the remaining Subfigure (2). This diagram is a scatterplot that confronts - like before in Figure 5.1, Subfigure (2) - the values for accuracy-error along the horizontal axis and the values for precision-error along the vertical axis.

Whereas this time, the values are the differences for these measures. A point in this diagram represents the differences in accuracy-error and precision-error from the first block to a certain later block for a certain participant. Based on these points, we also determined the non-dominated fronts which we will use in later analyses. Here, we show these fronts again as polylines connecting the corresponding points. To emphasize increase and decrease, or improvement and degradation, we added orthogonal to both axes lines that mark the border between both different developments at the value 0. This gives us 4 quadrants. The lower left quadrant shows us an area with improvement in both measures, the upper left improvement in accuracy-error and degradation in precision-error, the upper right degradation in both measures, and the lower right quadrant an area with degradation in accuracy-error and improvement in precision-error. We see that the points are distributed across all quadrants. So all combinations of developments occurred which is different from the empty triangle in the previous section. We also see points with no development in one of the measures or no development in both. Generally, the density of points near the intersection at (0, 0) is higher than in the distant parts. A tendency of improving regarding both measures, we see from the fact that the lower left quadrant for improvement regarding both measures contains the most points in comparison to the other quadrants. Regarding the non-dominated fronts, we see the best fronts starting from the lower left to the worst fronts in the top right. The best fronts span from the upper left quadrant over the lower left quadrant to the lower right quadrant. Doing so, they combine points with strong improvement in one and degradation in the other measure with points that show improvement in both measures. The worst front span inside the upper right quadrant combining points with degradations according to both performance measures.

So far, we examined the occurring values for development in performance measures that we will confront in Section 5.3.3 with development in distribution properties of our throw features in order to identify characteristics of participants that improve more than others. We observed the tendency to improve according to both measures. Apart from improvement, we also observed cases of no development and degradation. This gives us the possibility to confront cases of development later on. Different from the fronts on the actual values for the performance measures in the previous section, this time, especially the best fronts span a full range and are not just influenced by just one of the measures. Here they combine both accuracy-error and precision.

5.2 Relation between Performance and Throw Pattern

In the following, we want to analyze the performance data in their relation to the throw features. We expect that better participants show a higher variety in their throw physiology. So far, we determined for each of the 5 blocks - so for every 100 throws - of a participant the performance measures accuracy-error and precision-error, as well as the distribution properties median and IQR for each throw feature. Applying IQR to a throw features gives us an physiological variance measure. At this point, we have for each participant 5 data points which are the basis for the following Figures 5.4, 5.5, and 5.6. Each of them shows in Subfigure (0) a visualization of one or both performance measures as well as in Subfigures (1) to (20) the distribution measures for each feature. Looking at Figure 5.5, Subfigure (6) just as an example, we see a scatterplot that opposes for the data points of all participants for the feature d_t2_t4 the median along the horizontal axis as well as the IQR along the vertical axis. In addition, we see data points colored in transparent grey, blue or red. Yet in order to get to know this kind of diagram, we will ignore the meaning of the color for this moment. Considering the vertical axis on IQR, we observe that the elements of the blue group of data points tend to reach higher values for IQR than the elements of the red group. For the feature of this subfigure - d_t2_t4 - which is the latency between the time of the maximum acceleration and the time of maximum deceleration, this means that the variations of this latency for the blue group tend to be bigger than the respective variations for the red group. On the other hand considering the horizontal axis on the median, we see that the red group is more **concentrated** as it covers a smaller interval between values from 100 to 150. However, the blue group seems to be more **scattered** as it covers a greater interval. Regarding the latency between peak acceleration and peak deceleration this means that the elements of the red group tend to be more similar in their average latency. Whereas the elements of the blue group reach higher as well as lower average latency being less similar. Also, when we consider both axes, we see that the red points are more concentrated than the blue points showing that the elements of the red group are also more similar regarding IQR.

In order to find answers for our research question of which differences exist between participants with a high performance and participants with a low performance, we define a group of good elements with a high performance and a group of bad elements with a low performance based on different criteria regarding our performance measures - accuracy-error and precision-error. Having these groups, we oppose both of them visually to look for differences in their distribution properties for throw features. In the following Figures 5.4, 5.5, and 5.6, we show the good group in red color and the bad group in blue color. Elements that we do not assign to a group, we show in a transparent grey for context. Applied to our initial example from Figure 5.5, Subfigure (6), we see that it is the group of good elements in red that shows less variance in the given feature and whose elements are more similar to each other than the elements in the bad group in blue. To shorten the following descriptions, we abbreviate the good red group as **R** as well as the bad blue group as **B**.

According to our notion of variability in motor control, we expect participants with a higher performance to show higher variance in their throw physiology. We quantify this variance using the IQR distribution property of throw features over a whole block of throws. In consequence, we expect for the scatterplots for the distribution measures of the throw features to see that R shows higher IQR values than B. So what we see in our example in Subfigure (6) does not match to what we expect. It rather opposes what we expect as R shows less IQR instead of more IQR.

In the next three sections, we define R and B, first and second solely based on one of our performance measures and third based on both of them using fronts from non-dominated sorting. Each time, we use these groups to identify differences according to the throw features and check whether they fit our expectations. Afterward, we summarize and compare the sets of observations that we make.

5.2.1 Groups based on Accuracy-error

In this first analysis, we select R and B solely based on the accuracy-error performance measure. As this measure is an error measure it is better if the error's value is smaller. Because of this, we selected for R the 10 elements with the smallest values for accuracy-error. On the other hand, we selected for B the 10 elements with the highest values for accuracy-error. In Figure 5.4, Subfigure (0), we see a histogram on the occurring values of accuracy-error.

With respect to our selection R forms the left border with the smallest values while B forms the right border with the highest values.

Based on this definition of R and B from accuracy error, we will now describe our observations of differences between both groups regarding the distribution measures for each throw feature in Figure 5.4. Yet, we will not describe features or in other words, Subfigures in which we see no clear difference.

Regarding the features d_t2_t3 and d_t2_t4 in Subfigures (5) and (6), we observe that R except for two outliers is more concentrated according to IQR also tending to have smaller IQR than B. This opposes our expectations as we expect higher IQR for the elements of R.

Looking at Subfigure (13), we see R having the tendency to have higher IQR than B which fits what we expect.

Lastly, in Subfigure (19), we observe - ignoring two outliers - B tending to have lower IQR for a given median. So also here R shows what we expect, i.e., a higher IQR than B.

Using solely the accuracy-error to define R and B, we observed differences in 4 of 20 throw features according to their distribution properties. Though the differences for 2 features match our expectations there are also 2 features that show differences that oppose them.

5.2.2 Groups based on Precision-error

In this analysis, we define the selections for good as well as bad based solely on the precision-error performance measure. As this measure is an error measure like the accuracy-error in the previous analysis, we consider smaller values preferable or better than bigger values. Respectively, we selected for the group of good elements R the 10 elements with the lowest values in precision-error and for the group of bad elements B the 10 elements with the highest values. Looking at the histogram in Figure 5.5, Subfigure (0), we see R at the left border of the distribution with the lowest occurring values and we see B at the right border with the highest occurring values.

Regarding the Subfigure (1) for the first feature d_t1_t2, we see R and B separated by a conceivable diagonal. So for a given median R shows smaller

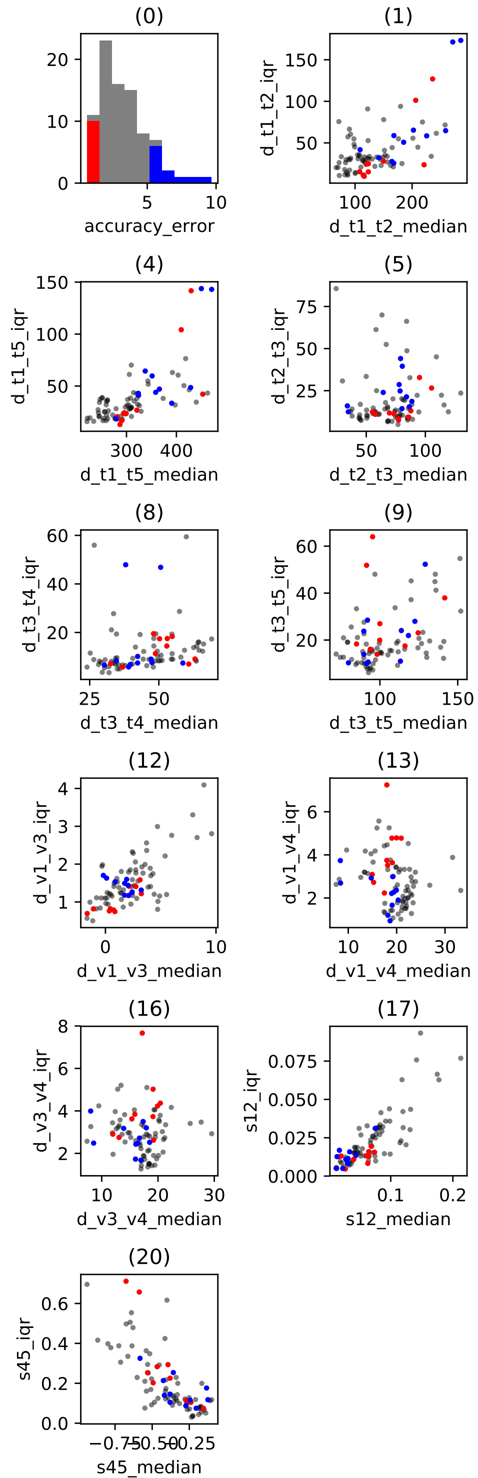

Figure 5.4: Confrontation based on accuracy-error - Subfigure (0) represents the selected good and bad group based on accuracy-error. Subfigures (1) to (20) show the distribution properties of these groups for each throw feature.

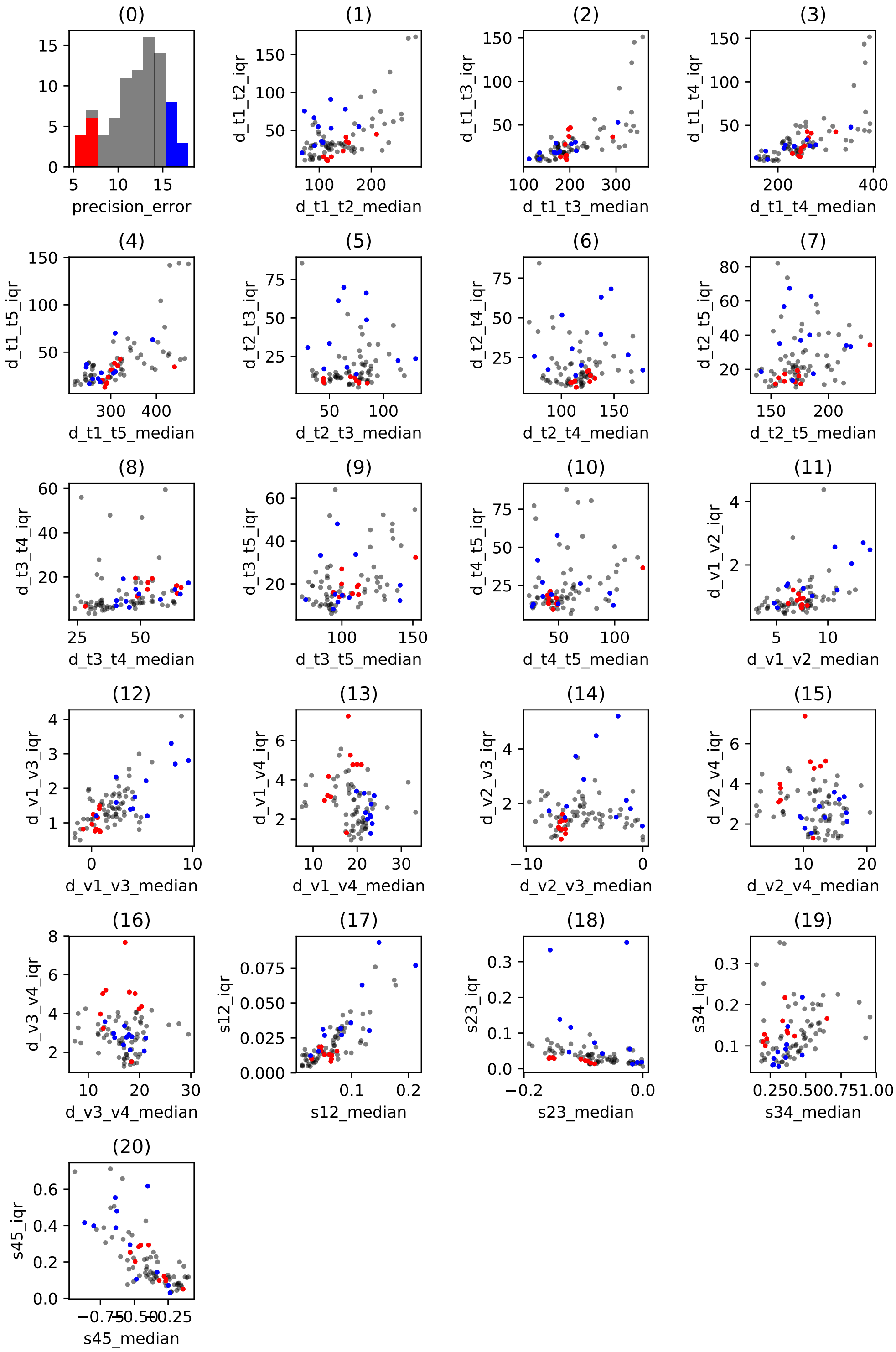

Figure 5.5: Confrontation based on precision-error

values for IQR than B which opposes our expectation of R having higher IQR values.

Looking at Subfigure (5), we observe R having lower values and being less scattered according to IQR. This stands again against what we expect.

Similar to this, we see in Subfigure (6) the same properties for R. In addition R is also more concentrated regarding the median values.

In Subfigure (11) for d_v1_v2 which represents the maximum acceleration R and B intersect. Yet R is more concentrated with small values for median and IQR. B tends to have higher IQR, which is not what we expect.

In the next subfigure - Subfigure (12) -, we see a similar yet stronger difference as R and B just slightly overlap while R reaches lower values for both distribution measures.

Next, we look at Subfigures (13) and (15) for the related features d_v1_v4 and d_v2_v4 which represent respectively the maximum deceleration and the difference between maximum acceleration and maximum deceleration. Both times, we observe - ignoring one outlier each time - that R tends to have higher IQR and smaller median. This fits both times our expectation.

Regarding d_v2_v3 in Subfigure (14), we observe R to be more concentrated according to IQR as well as median. B reaches higher values for the median as well as the IQR. This opposes what we expect to see.

Looking at Subfigure (17) for s12 - the slope at the beginning of the acceleration phase -, we observe R and B intersecting yet R being concentrated with small values in both distribution measures. On the other hand, B reaches higher values for median and IQR which opposes our expectations.

Lastly, we look at s23 in Subfigure (18) which opposes again what we expected as R shows lower IQR for a given median.

Using solely precision-error to define a good group and a bad group, we observed differences between those groups in the distribution measures for 10 of 20 throw features. Here just 2 observations match our expectation of R having higher IQR while 8 observations show the opposite.

5.2.3 Groups based on Fronts from non-dominated Sorting

As we are interested if we get more insight by not looking at a single performance criterion but instead considering multiple performance measures in a combined way, we determine our groups in this last comparison based on the fronts from non-dominated sorting on both performance measures - precision-error and accuracy-error.

As these are error measures smaller values are more preferable. So smaller values are more dominant for both measures during the non-dominated sorting. We see the resulting fronts in Figure 5.6 Subfigure (0) as polylines with the best fronts starting from the left going to the worst fronts in the upper right.

Having these fronts, we define R - the set of good elements - using the best three fronts which contain in total 14 data points. For B - the set of bad elements -, we select the last four fronts which contain a total of 15 data points.

In Subfigure (0), we see R and B respectively in red and blue. R contains the elements with the lowest values for accuracy-error and spans from elements with the lowest occurring values for precision-error almost up to elements with the highest values. This is a consequence of the absence of data points with low precision-error and high accuracy-error which we discussed in Section 5.1. So R consists mainly of elements from a slice starting from the left solely influenced by accuracy-error. On the other hand, B consists of elements with a medium to high accuracy-error and precision-error.

Looking at Subfigures (5), we see for d_t2_t3 as well as d_t2_t4 which is the latency between the time of maximum acceleration and the time of maximum deceleration that R tends to be more concentrated with lower values for IQR. This opposes our expectations.

Also, regarding d_v1_v2 which is the maximum acceleration, we see in Subfigure (11) that R is more concentrated regarding IQR. Here it is intersecting with the lower part of B. Yet again it opposes what we expect to observe.

In Subfigure (12), we observe that R tends to have lower values for median and IQR than B. This is contrary to our expectation.

Regarding Subfigures (13) and (15), we observe the distribution measures for d_v1_v4 and d_v2_v4 respectively the maximum deceleration and the difference from maximum acceleration to maximum deceleration. Here R tends

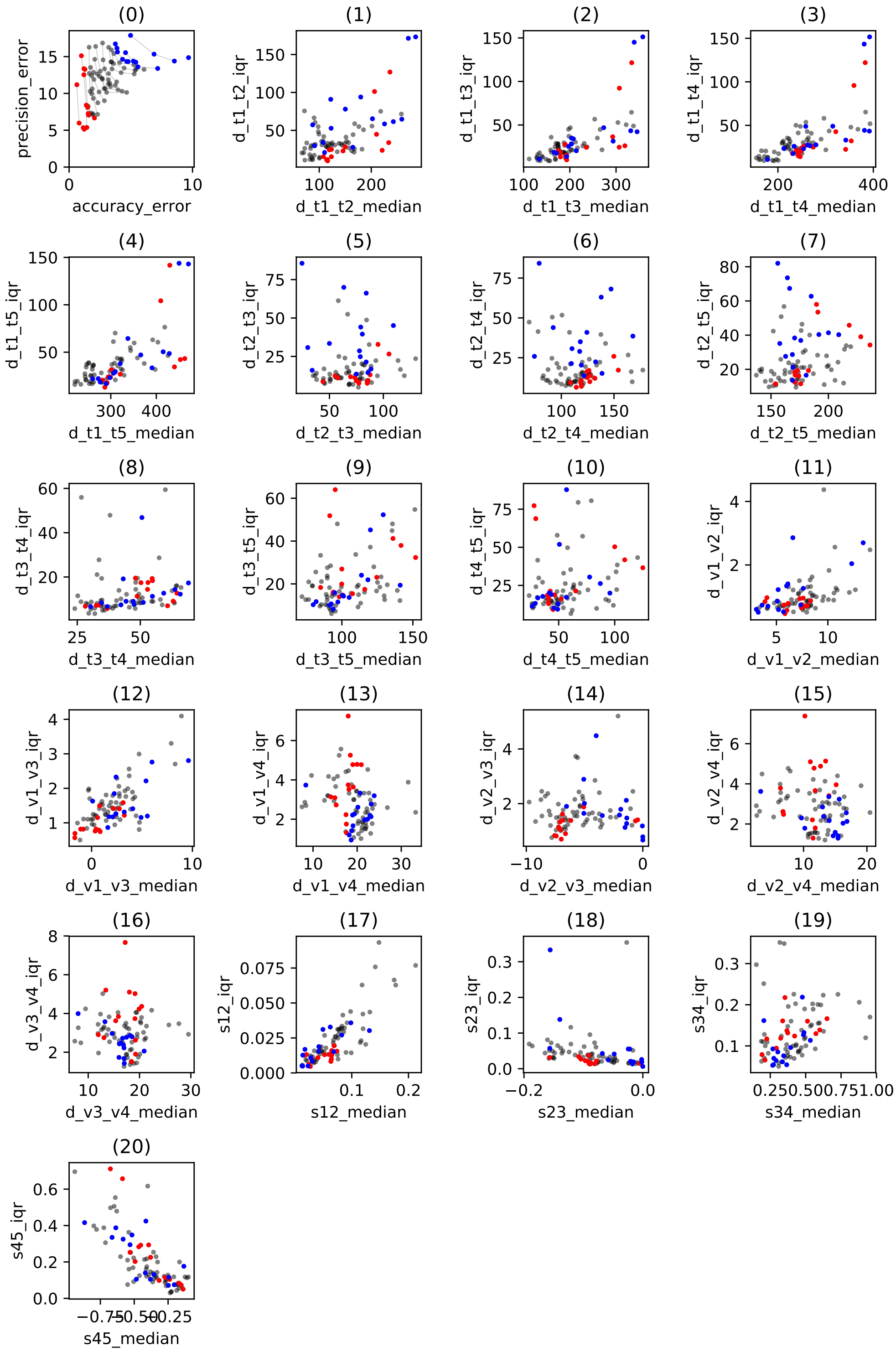

Figure 5.6: Confrontation based on non-dominated fronts on both performance measures

to be smaller regarding the median but it is also more scattered regarding IQR and tends to reach higher values in it. This matches what we expect.

Also in Subfigure (16) for d_v3_v4, we see R being more scattered and tending to reach higher values for IQR which fits our expectations.

For s12 - the slope towards the maximum acceleration - in Subfigure (17) R is more concentrated with low values according to both distribution measures which opposes what we expect.

Lastly regarding Subfigure (18) for s23 which is the slope of decreasing acceleration after maximum acceleration R tends to have smaller median than B. This neither fits nor opposes our expectations.

Using groups based on the fronts from non-dominated on our performance measures, we found differences in the distribution measures of 9 of 20 throw features. Though 3 observed differences match our expectation of higher IQR for R there are also 5 observations that oppose it.

5.2.4 Comparison of Sets of Observations

In each of the previous three sections, we defined a good and a bad group based on different strategies. The first time, we defined the groups based on the accuracy-error, the second time based on precision-error, and the third time based on the fronts from non-dominated sorting to combine both measures. We used these groups to identify throw features that showed differences in the distribution measures according to these groups.

In total, we made 23 observations of differences which we show in Table 5.1. There, we marked that we observed a difference for a certain feature using a certain strategy to define the groups by marking the respective table cell with an "x".

To get more insight into the data, we decided at the beginning of the development of our method to use the two performance measures accuracy-error and precision-error, and to combine them using non-dominated sorting. Now, we see in the table that the accuracy-error based strategy gave us - with 4 observed differences - the fewest observations. The other two strategies gave us - with about 10 each - more observations. Yet, for each strategy, the set of observations is different to the others and also contains at least one observation that is unique as it appears under no other strategy.

In Section 5.2.3, we described that the selected R for the front-based selection resembles a slice of data points according to just accuracy-error which we can see in Figure 5.6, Subfigure (0). Also, B appears to be close to being selected by another slice just on accuracy-error except for some data points below B. With this impression the groups for that section are very similar to the groups from Section 5.2.1 with groups exclusively based on accuracy-error which are represented in Figure 5.4, Subfigure (0). Nevertheless, in Table 5.1, we see that for the front-based selection that we have more than double the amount of observed differences making it very different than the accuracy-based approach. According to the number of observations, the front-based approach is more similar to the precision-based approach.

Regarding our expectations of higher variance for better players, we assessed for each feature in which we saw a difference between the good and the bad group whether this difference matches to our expectation or whether it opposes showing the opposite. We summarized also this in Table 5.1 by marking a match with a "(+)" as well an opposition with a "(-)". 22 of 23 observations are either matching or opposing what we expect. Only one - for the feature s23 in the front-based approach - shows a difference that neither matches nor opposes as the difference relates exclusively to the median which we see in Figure 5.6, Subfigure (18). We want to emphasize that our expectations just focus on IQR and we also observed differences in median in many other cases. Looking row-wise through the table comparing the matching of observations for a feature using different approaches, we see for 7 of 8 rows with multiple observations that they show the same whether they match or oppose. Overall, we made 7 observations that fit our expectation but also 15 that oppose.

5.3 Relation between Performance Developments and Throw Pattern Developments

Based on our definition of development as the difference from the first block to the later blocks, we have 4 data points for each participant. Each data

Table 5.1: Overview of throw features with observed differences in performance - We defined a good group and a bad group according to different strategies. When we observed a difference between the groups according to a certain strategy (column) in a certain feature (row) then we mark it here with a **x**. In addition, we marked differences that fit or oppose our expectations respectively with **(+)** or **(-)**.

subfigure	feature	accuracy-based see section 5.2.1	precision-based see section 5.2.2	front-based see section 5.2.3
1	d_t1_t2		x (-)	
5	d_t2_t3	x (-)	x (-)	x (-)
6	d_t2_t4	x (-)	x (-)	x (-)
11	d_v1_v2		x (-)	x (-)
12	d_v1_v3		x (-)	x (-)
13	d_v1_v4	x (+)	x (+)	x (+)
14	d_v2_v3		x (-)	
15	d_v2_v4		x (+)	x (+)
16	d_v3_v4			x (+)
17	s12		x (-)	x (-)
18	s23		x (-)	x
19	s34	x (+)		

point consists of differences for performance measures as well as differences of distribution measures for throw features. In the following Figures 5.7, 5.8, and 5.9, we show those differences for performance measures each time as Subfigure (0) and for distribution measures in Subfigures (1) to (20).

Looking at Figure 5.7, Subfigure (1) as an example, we see a scatterplot that confronts the differences of median as well as IQR for the feature d_t1_t2. In order to understand what we see in this type of diagram, we look at the set of blue elements and ignore the meaning of this set for now. According to median that is represented along the horizontal axis, we see that this blue set shows small as well as above yet positive difference values. Some lay with a very small difference value close to the vertical 0-line showing very small or no development. Others, we see further to the right showing us an **increase** in the median distribution measure. For the feature d_t1_t2 that captures the time from the start of the acceleration until the maximum acceleration, this means that the elements of the blue group take longer on average later on to reach that peak acceleration. Switching to the vertical axis for the differences in IQR, we see for some elements of the blue group that they are very close to the horizontal 0-line or in other words that their difference value is very close to 0. These show again very small or no development. The rest of the blue group shows absolute bigger but negative difference values in IQR for the given feature which makes them represent a **decrease** in IQR. For the d_t1_t2 feature, this means that the variance of the time to the maximum acceleration got smaller for these elements, which may mean that the process of throwing got more stable or similar.

We see in this blue group for the d_t1_t2 feature a tendency for an increase in median with a decrease in IQR though some elements show no development in median or IQR. This observation may be more significant or convincing if all elements would show this increase and decrease. Yet the blue group is more clear in its development being more **concentrated** compared to the red group. While some elements of the red group lay close to the blue elements there are others that actually show stronger developments but also opposite developments which makes it more diverse and appear more **scattered**. For example, there are also red elements in the bottom right of the figure that show a decrease in median as well as IQR.

Returning to our research question of what are differences between participants which improve a lot and participants which do not we will define in the following a group that is good or better according to a chosen criteria based on the development as well as another group that is bad or worse based on the same criteria. We show the good group as the red group as well as we show the bad group as the blue group. For our example from before from Figure 5.7, Subfigure (1) this means that both groups tend to reduce IQR. Yet they are different in the development according to median as the bad group just shows an increase while the good group shows a decrease as well as increase.

According to our idea of variability in motor control, we expect that the improvement of participants is related to an increase in variability which is here represented by the distribution measure IQR for throw features. According to this, our good group should show positive IQR difference values which represent an increase in IQR. Comparing both groups, we expect the IQR differences for the good group to be bigger than the IQR differences for the bad group. Considering our example from before, we see that the red good group and the blue bad group are not separated according to IQR differences. Also, for both, we rather see a decrease in IQR than an increase in IQR. Though some elements show an increase in IQR yet our example shows a combination that we did not expect.

In the following three sections, we observe differences in the developments of throw features of our good and bad group based on three different strategies to define those groups. First and second, we define the good and bad group solely based on the development in one of our performance measures. Third, we define them based on fronts from non-dominated sorting on developments in both measures. Afterward, we compare the sets of observations that we make.

5.3.1 Groups based on Accuracy-error

Now, we want to define both groups based on the development in the accuracy-error performance using differences in that measure from the first block to later blocks. The good group is again assigned with the color red. We abbreviate it in the following observations with **R**. On the other side the bad group is assigned again to the color blue. So, we abbreviate it in the following observations with **B**. We selected the 11 elements with the biggest negative difference values as R. They represent the elements with the biggest improvement as their accuracy-error decreased the most from the first to the later block. In addition, we selected the 11 elements with the biggest difference values as B. Looking at Figure 5.7 Subfigure (0), we see that they show a degradation as their accuracy-error increased. From here, we can see that R and B represent the left and right end of the distribution of occurring difference values.

Let's describe the observations for differences according to a good and a bad group defined based on the development in accuracy-error performance measure. We focus exclusively on Figure 5.7. Yet, we do not describe subfigures in which we cannot see a clear difference between the groups.

Looking at Subfigure (1) for the development of the distribution measures of the d_t1_t2 throw feature, we see that B as well as R mostly change a little or rather gets smaller in IQR. This does not match our expectations. However, we see a difference in median as R is more scattered than B showing partially an increase and partially a decrease in median while B shows just an increase in median.

Looking at Subfigure (2) for the development of the distribution of d_t1_t3, we see that B is more concentrated with the tendency of increasing the median while R is more scattered with different combinations of increases or decreases in median and IQR. We see similar developments for d_t1_t4 in Subfigure (3) as well as d_t1_t5 in Subfigure (4).

In Subfigure (6) for d_t2_t4, we see for B a slight decrease in median time whereas R shows rather a decrease in IQR. This means the latency between peak acceleration and peak deacceleration gets smaller for B whereas for R it is more characteristic that the variance of that latency decreases meaning that the throws getting more similar regarding that feature.

Regarding Subfigure (8) for d_t3_t4 B seems to be slightly more scattered according to developments in IQR. Yet, we see no difference that matches our expectations.

In Subfigure (9), we see for the total time of deceleration - d_t3_t5 - for R mainly an increase in IQR whereas B shows rather a decrease in IQR. This matches our expectations.

Starting to look at the amplitude features, we see for the maximum acceleration - d_v1_v2 - that B decreases in IQR while R shows smaller also increasing developments. Also, this difference does not support our expectations.

In Subfigure (15) for d_v2_v4 - the amplitude difference between the maximum acceleration and the maximum deceleration -, we see no or in comparison

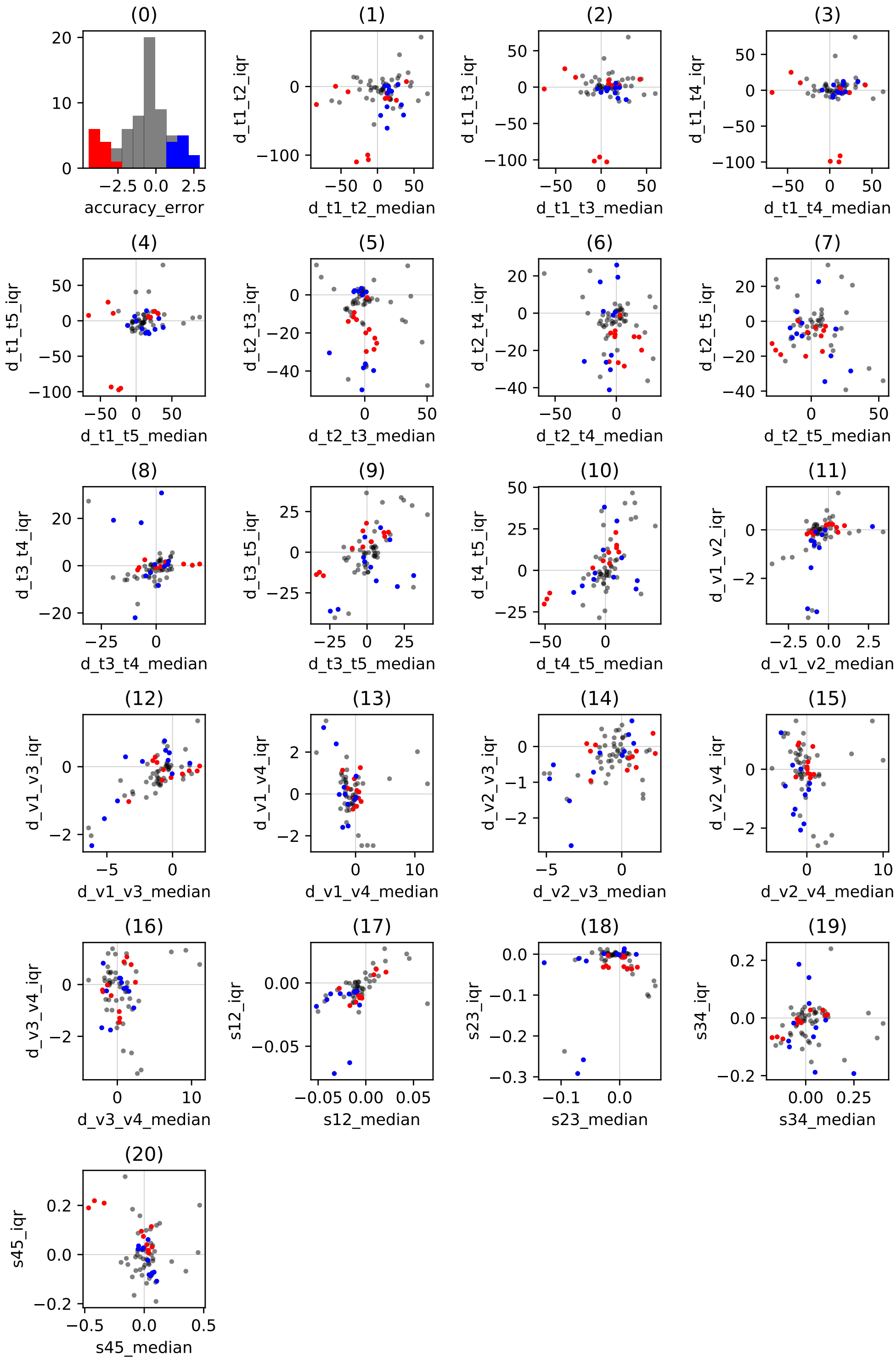

Figure 5.7: Confrontation based on development of accuracy-error

to the rest small changes for R. On the other hand, B shows a tendency for a reduction in median while it is also more diverse in IQR. Also this difference, we cannot fit to our expected differences.

Looking at Subfigure (18) for s23, we see as the only difference between B and R that R is less scattered than B which we cannot fit to our expectations.

For Subfigure (20) the feature s45, we see just for R an increase in IQR. This fits our expectations.

Summing up, we found 11 features with differences in development for both groups based on the developments in accuracy-error, e.g., different developments in median or IQR, or differences in the variations of the strength of developments. Yet, there were just 2 cases that actually fit our expectations on increasing variance.

5.3.2 Groups based on Precision-error

This time, we define the groups based on the developments in precision-error that is defined as the differences in precision-error of the first block to later blocks. Like before, we select for the red good group - **R** - the 11 elements with the smallest difference values. As we can see in Figure 5.8, Subfigure (0) these are elements at the lower border of the distribution of occurring differences with a negative difference that stands for a decrease in precision-error. So they are the elements that improved the most according to the precision-error. On the other side, we assigned the 11 elements with the biggest difference values which are closest to the upper limit of the distribution in Subfigure (0) to the blue bad group - **B**. They show differences close to 0 or positive. So they show a small to medium increase in precision-error which is a degradation.

Looking at Subfigure (1), we see for B a tendency to slightly increase in median as well as to slightly decrease in IQR. R is different as it shows stronger developments that are also more diverse because there are in addition elements with opposite developments as described for B.

In Subfigure (5) R appears again scattered compared to B without a certain tendency for a common development. On the other hand, B is rather concentrated with a tendency to slightly decrease in median.

In Subfigures (6), (7), (9) as well as (20), we see B as rather concentrated while we see R rather scattered without a certain tendency for the developments.

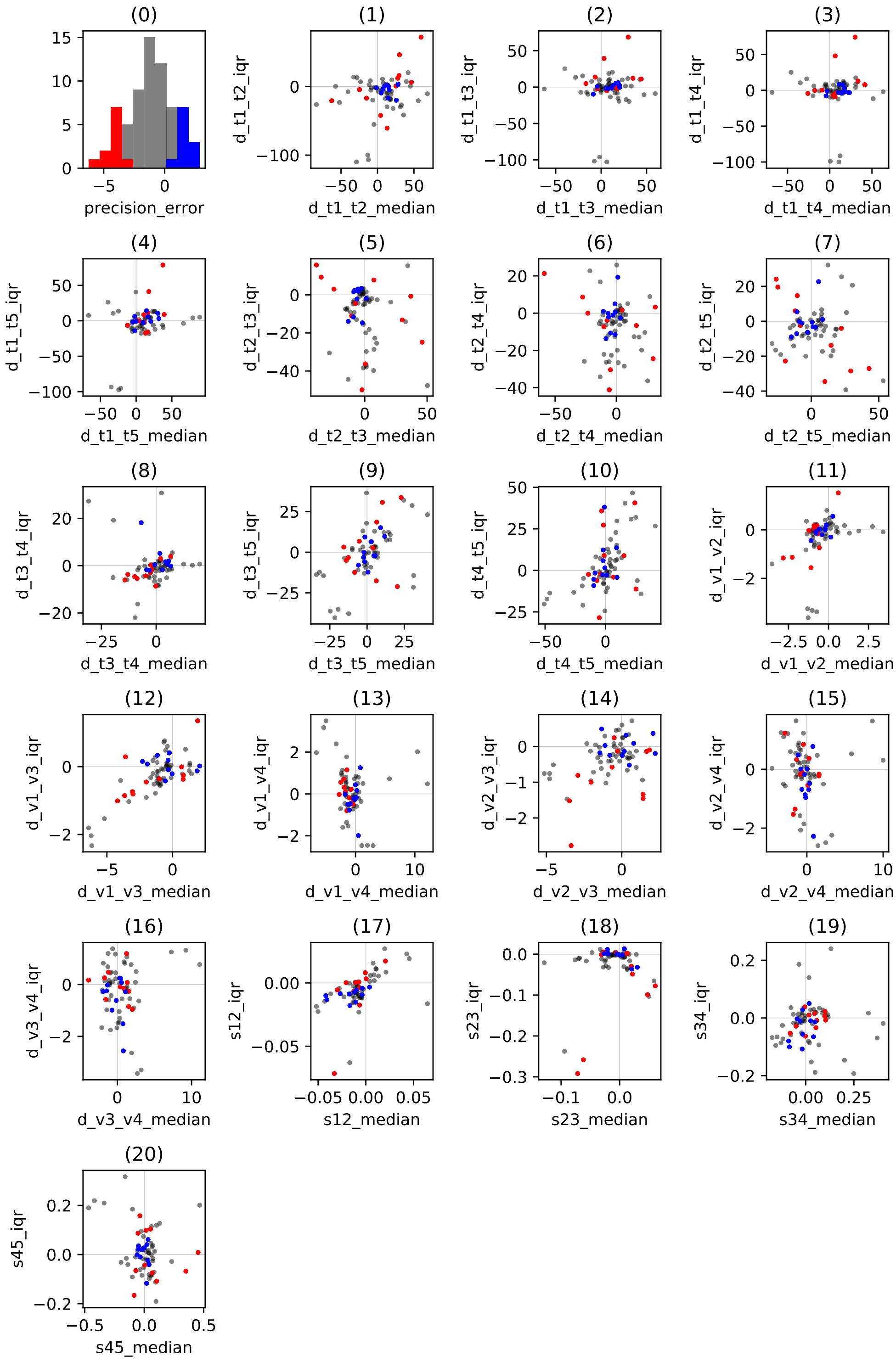

Figure 5.8: Confrontation based on development of precision-error

Regarding d_v1_v4 in Subfigure (13), we observe a stronger tendency for R to reduce in median compared to B.

Looking at d_v2_v3 in Subfigure (14), we see for R decrease in IQR that we do not see for B. This is opposing what we expect.

In Subfigure (17) for s12 which represents the slope towards the maximum acceleration, we see for B the tendency to decrease in median as well as in IQR. We do not see this tendency for R.

Summing up, we identified 9 features with differences in development based on groups from development in precision-error. None of these observations actually fits our expectations while one of them even opposes our expectations.

5.3.3 Groups based on Fronts from non-dominated Sorting

In this analysis, we define R and B based on fronts from non-dominated sorting. Similar to the application of non-dominated sorting in the analysis of status for the multi-criteria case in Section 5.2.3, we determine the fronts based on minimizing each measure. In that analysis, a smaller value for each of the performance measures was better as they represent errors where a smaller error is preferable. This time, we use the developments instead, i.e., the differences in the performance measures from the first to later blocks. Yet, also with these, a smaller value is preferable as we consider differences now. Negative values represent a decrease in the errors while positive differences represent an increase in error or in other words a degradation. The result of the non-dominated sorting, we see in Figure 5.9, Subfigure (0) where the resulting fronts are visible as polylines with the best fronts starting in the bottom left and the worst fronts in the top right. For the group of good elements that we show again in red and abbreviate with **R**, we selected the elements of the first 2 fronts with 12 elements. So the group size is close to the group size in the previous analyses. As we can see in Subfigure (0) R contains elements with strong improvements in both performance measures showing elements with negative difference values in both performance measures. Yet, at the ends of the selected fronts, we see elements with strong improvement in one measure while there is a degradation in the other measure. For the group of bad elements - **B** -we selected the elements of the last 4 fronts with 10 elements.

In Subfigure (0), we see that this group consists of elements that slightly or stronger degrade in both performance measures.

Looking at Subfigure (1), we observe B being more concentrated with a slight increase in median while R shows increases as well decreases in the median of the feature d_t1_t2. In addition for R, we see a tendency to reduce in IQR. This, plus the fact that B shows almost no development in IQR seems to show the opposite of what we expect.

Regarding the next feature d_t1_t3 in Subfigure (2), we also see that B is more concentrated with a tendency to slightly increase in median while R shows stronger developments with increases as well as decreases in median. This also applies for the following feature in Subfigure (3).

In Subfigure (5) for d_t2_t3, we observe that B shows no development in IQR while R shows decrease in IQR. This observation opposes our expectation.

Regarding the next Subfigures (6) and (7), we see a similar difference for B and R like in Subfigure (5). Again R shows a stronger decrease in IQR which opposes our expectation.

Looking at Subfigure (8) for feature d_t3_t4, we observe for R a tendency to decrease in in median while for B we observe a tendency to increase.

In Subfigure (12), for B as well as R, we see a tendency to decrease in median. Yet R decreases stronger.

For feature d_v2_v4 in Subfigure (15), we observe for B the tendency to decrease in IQR while for R we see increases as well as decreases in IQR. This fits partly to our expectations. Though the B and parts of R show decrease in IQR nevertheless other parts of R show the expected increase in IQR and the stronger increase than B.

In Subfigure (16) for d_v3_v4, we observe for R the tendency to increase in median while in B we see a weaker increase and also decrease.

Lastly, in Subfigure (18), we see again for R a stronger tendency to decrease in IQR while B is rather concentrated with no development in IQR which opposes our expectations.

Summing it up using groups based on fronts from non-dominated sorting, we found 11 features that showed differences in development according to those groups. We found no feature in which the groups showed the development that

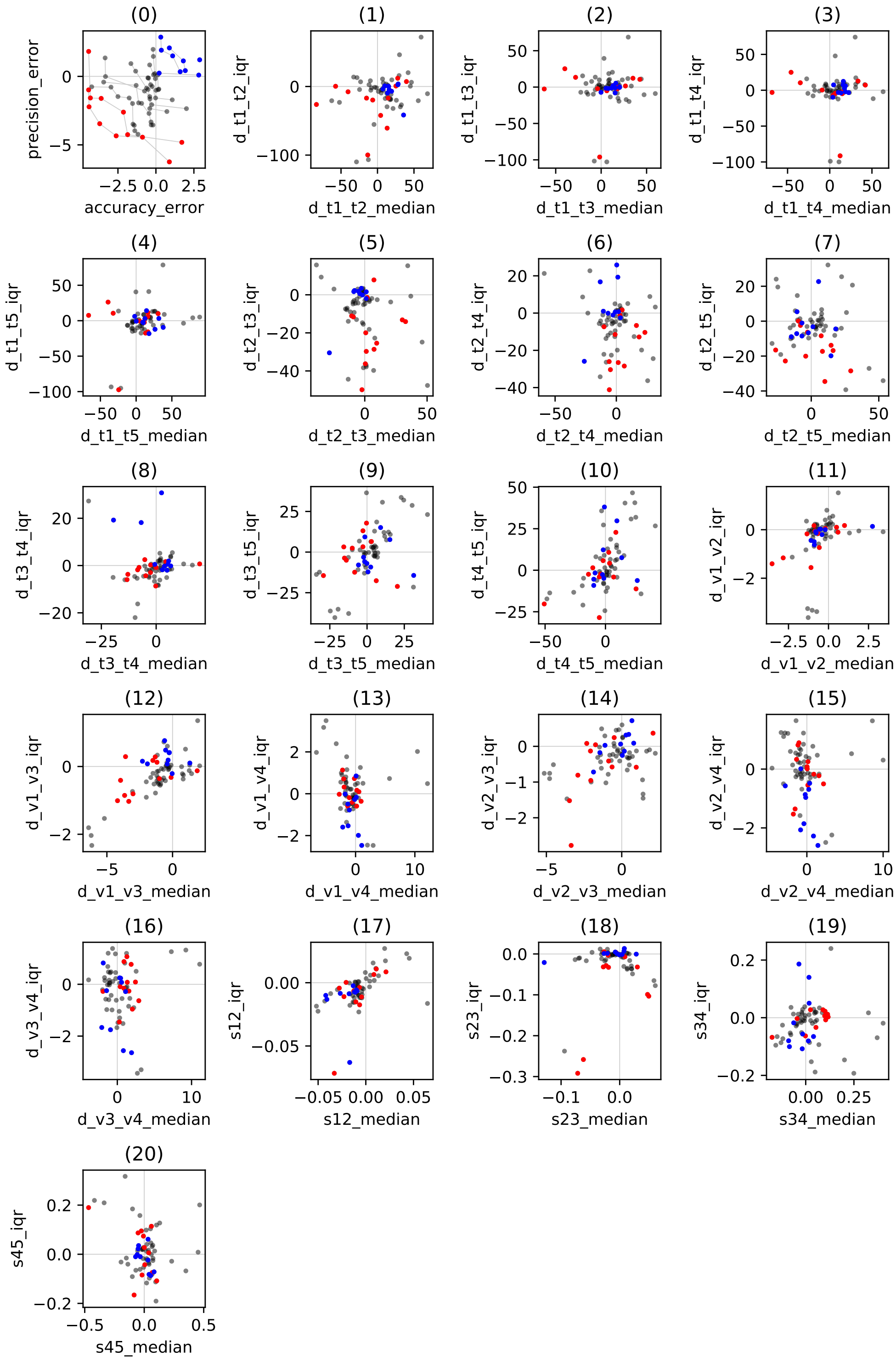

Figure 5.9: Confrontation based on fronts from non-dominated sorting on the development in accuracy-error and precision-error

we expected. Yet, we found 5 features where the groups showed developments that rather oppose our expectations.

5.3.4 Comparison of Sets of Observations

We made 3 analyses using 3 different strategies to define good and bad groups. For the first two analyses, we used the development in one of the performance measures each. For the last analysis, we combined the development in both measures using the fronts from non-dominated sorting. We opposed these groups in each analysis in order to find differences in the development of features that fit our expectation of a rising variance.

As we summarize in Table 5.2, we found 31 differences for 18 of 20 features, in at least 9 features in each of our analyses.

Though we see different developments many times, most of the time those observed differences do not match to our expectation. Even 5 times they seem to oppose our expectation and there are just 2 observations that match to them.

Our initial motivation to use fronts from non-dominated sorting to combine both performance measures was to observe effects that could not be found using single performance measures. After these 3 analyses, we see that the selected groups for the 3rd analysis - that we can see in Figure 5.9, Subfigure (0) - could not have been selected by the single measure approach on this data. For R, we combined elements with the smallest values according to accuracy error which were selected for R in our first analyses with elements with the smallest values according to precision error which were selected for R in the second analyses. For B it is more like an intersection, i.e., selecting elements that just appear in both sets, of the Bs that we selected in the first two analyses. Also, the observations that we made appear to us like a combination of both measures as for 12 of 16 cases, we see in Table 5.2 that if we observed a difference in the 1st or the 2nd analyses then we also observed a difference in the 3rd analyses. Yet, what we observed in the 3rd analysis is also different because we see the most cases of observed differences that oppose our expectations. In addition for 2 features, we saw differences that we did not see in the analyses that were based on the single performance measures.

Table 5.2: Overview of throw features with observed differences in development - We defined a good and a bad group according to different strategies. When we observed a difference between the groups according to a certain strategy (column) in a certain feature (row) then we mark it here with a **x**. In addition, we marked differences that fit or oppose our expectations respectively with **(+)** or **(-)**.

subfigure	feature	accuracy-based see section 5.3.1	precision-based see section 5.3.2	front-based see section 5.3.3
1	d_t1_t2	x	x	x (-)
2	d_t1_t3	x		x
3	d_t1_t4	x		x
4	d_t1_t5	x		
5	d_t2_t3		x	x (-)
6	d_t2_t4	x	x	x (-)
7	d_t2_t5		x	x (-)
8	d_t3_t4	x		x
9	d_t3_t5	x (+)	x	
11	d_v1_v2	x		
12	d_v1_v3			x
13	d_v1_v4		x	
14	d_v2_v3		x (-)	
15	d_v2_v4	x		x
16	d_v3_v4			x
17	s12		x	
18	s23	x		x (-)
20	s45	x (+)	x	

6 Discussion

In this chapter, we will first interpret whether our method and results fulfill our goals and expectations. Afterward, we will give answers to our research questions. At last, we will show ideas for future work to improve and extend this work, and to find answers to related questions.

First, we like to recap this thesis. In Chapter 1, we said that we need to learn movements for everyday activities. Based on the example of playing darts, we planned to explore the relation between performance and the way of throwing and named our research questions accordingly. What are the differences between better and worse players? What are the differences between players that improve more than others?

In Chapter 2, we followed the concept of variety destroying variety. Based on this, we expected that better players show a bigger variety in their throws and that players that improve more also show a bigger increase in the variety of their throws. In addition, we decided to use multiple complementary measures to quantify performance because we expected to make additional findings.

To test our expectations, we wanted to quantify throw performance and throw physiology, as well as their developments over time, and to oppose them in order to explore their relation. Additionally, we wanted to combine multiple performance measures using non-dominated sorting in our analyses.

In Chapter 4, we proposed accuracy-error and precision-error to quantify performance. For quantifying the physiology of throwing, we proposed to describe the pattern of single throws regarding the time series of the norm of accelerating forces using throw features and to describe the variations of this pattern for a set of throws by determining median and IQR for every throw features (cf. Figure 6.1). Applying those measures to experimental data, we identified in Chapter 5 differences in the distribution properties of throw features between good and bad groups based the performance measures.

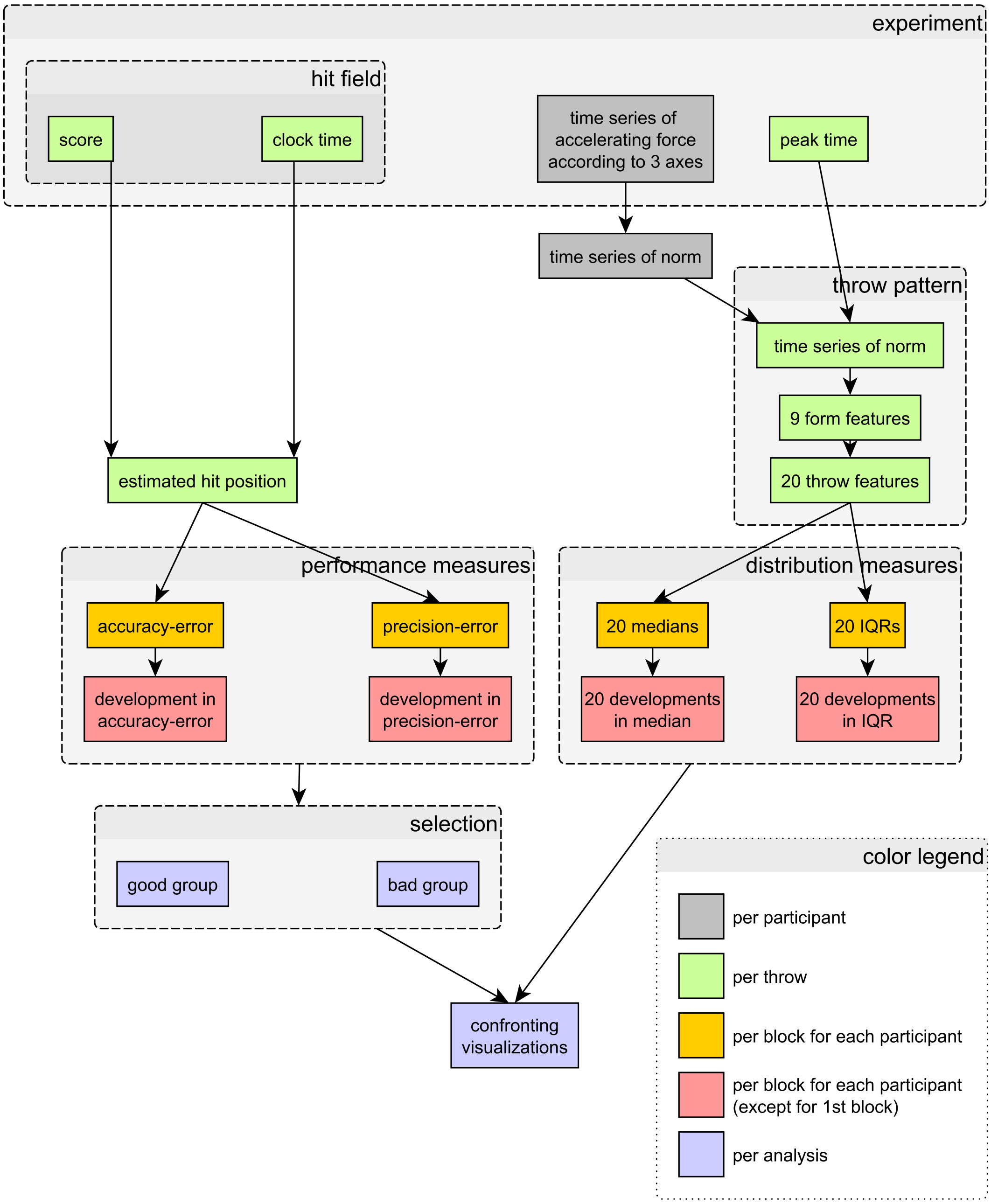

Figure 6.1: Overview of our method to prepare the confronting visualizations

6.1 Our Results

In this section, we will evaluate how our results match our expectations in order to give an answer to our research questions. We expected to make additional findings by using multiple performance measures and by combining them using fronts from non-dominated sorting. In Chapter 5, we found an unexpected triangle-like relation between our performance measures. In addition, we observed differences regarding the status of physiological measures and their development that were only visible with this approach. We expected to observe for better performing participants a higher variance in their throw physiology. Yet, in Section 5.2, we saw that much more observed differences were opposing this expectation showing better players to have a lower variance. Just for a few features that are related to the maximum deceleration - namely d_v1_v4, d_v2_v4, d_v3_v4, and s34 -, we observed differences that match those expectations. For stronger improving participants, we expected to observe a higher increase in variance of throw physiology. Again, the observations that we made in section 5.3 majorly did not fulfill this expectation. Mostly they did neither support nor oppose it. So generally, more of our observations opposed those expectations regarding the variance in throw kinematics. Apart from the influences of our method that we discuss in the next section, we see our expectations as a reason. They base on an understanding in which we assume that the variance in throw physiology is exclusively countering perturbations. This way, we ignored variance in the way of throwing that is not related to improvements in performance. We want to give an illustrating example. When we start shaking a participant who is throwing darts then the variance in throwing will increase but we do not expect that the performance will also improve. Quite the contrary, we expect the performance to degrade due to this additional perturbations. This leads us back to concept of [Latash2008] of good and bad variance. In this thesis, we were ignoring that certain variance in throwing decreases performance which is also part of the variance of the throwing physiology that we measure. To recap, Latash's good and bad variance describe relations between the variety of a movement and the variety of fulfilling its goal. Here, good variance represents a negative correlation as more variance in movement leads to less variance in fulfilling the goal. On the other hand, bad variance represents a positive correlation as more variance in movement leads to more variance in fulfilling the goal. [Boenke2018] subclassified Latash's bad variance. They name the relation of good variance quality variability. The relation of bad variance is separated into two types. Trivial variability is related to pursuing different goals. If a participant would alternatingly throw a dart towards a dartboard on a wall and drop a dart on a dartboard in front of their feet then we would observe more variety in their movement considering both movements a throw. The second part of bad variance is noise which relates to unintended disturbances of the movement. Shaking a participant during a throw increases the variance of the throwing but also decreases the performance. When we try to explore those correlations, we can measure the variance of the movement as we did in this thesis. Yet, this variance is compound from the variance of all three influences. As also the strength of these influences is unknown, the influences of triavial and noise variability may have superimposed the influences of quality variability. So, with our method, we may have observed the effect of trivial and noise variability instead of the influence of quality variability that we wanted to observe.

Based on the results of this thesis, we can answer our research questions like this.

- What are the differences between good and bad dart players?
  - For the participants in our experiment, we saw that better participants generally showed less variance in the kinematic properties of their throws.
- What are differences between participants that improved more and participants that improved less?
  - For the participants in our experiment, we saw mostly no clear difference according to the development in the variance of kinematic properties of their throws.

6.2 Our Method

In this section, we will evaluate whether our method fulfills our goals and point out in which parts we see strengths and weaknesses in our approach.

We will briefly summarize our method, first. We started with a dart throw experiment which we used to collect data for the physiology of throws and their outcome. Based on this, we quantified - on the one hand - performance according to two measures. On the other hand, we quantified physiology by focussing on the norm of the accelerating forces from the kinematic data, by fitting an acceleration model in order to extract throw features, by quantifying distribution properties of those features using IQR and median. We quantified the development in performance measures and physiological measures using the differences from the first to later blocks. We opposed performance and physiology by selecting good and bad groups according to single or both performance measure and visualizing these groups for every throw feature. Similarly, we opposed development in performance and development in physiological measures by selecting such groups and visualizing them regarding the distribution measures of throw features. Based on these visualizations, we identified differences regarding our expectations.

We see factors in our method that may influence the results to be less reliable and less convincing.

First, we want to emphasize the influence of the variance of participants from the experiment. We aggregate the throws block-wise to compute our performance and physiological-related measures. Then, we analyze the compound dataset including all 5 blocks for the status analyses and the 4 differences for each participant to later blocks in the analyses of the development. In Figure 6.2, we visualized again the values for accuracy-error and precision-error as we already did before in Figure 5.1, Subfigure (2) and in Figure 5.6, Subfigure (0). Yet, this time, we also connected for every participant the data points in the order of the blocks to a polyline. We see in the lower left of the diagram the polylines for three participants with low values according to both measures. They do not intersect each other and lay separated from the polylines for the rest of the participants. Comparing those three to the other polylines, we see that their points lay closer to each other. This matches the concept of the law of practice which states that under practice the performance changes less and less over time (cf. [Edwards2011]). The three advanced participants - with presumably more practice in their lives - show smaller changes in the performance measures between blocks. The participants with worse performance measure values - presumably beginners - show bigger changes in comparison of subsequent blocks which makes the points of their polylines lay further apart. Based on these measures, we select in our method the good and bad groups for opposing physiological throw properties visually. Since the points of those advanced participants lay closer together, it is more probable that more points

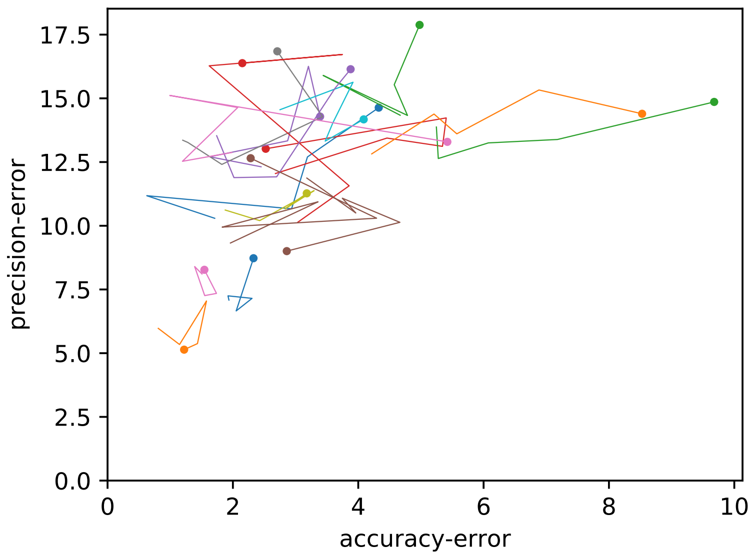

Figure 6.2: Traces of block performance - After aggregating the performance measures for every block of every participant we connected for every participant the data points in block order to a polyline. The start and respectively Block 1 is marked with a dot.

of the same participant get selected for a good group. As the points for the beginners are more scattered, it is more probable that the points of many different participants get selected for the bad groups. To exaggerate, we could imagine that for the good group we select only the points of a single participant and compare this group in our analyses to a bad group where each point belongs to a different participant. So, to a certain extent, we generalize for the bad performances more than for the good performances.

Furthermore, we see in Figure 6.2 that the majority of participants does not reach the proximity of the initial performance of the participants in the lower left. So, there are at least two classes of performance levels in the data in our experiment. This leads us to the open question of whether our participants are actually comparable using our method.

We see as another influence the errors that we introduce in our method. We have measuring errors during the recording of the experiment data. We introduce errors while fitting the model due to the nature of fitting a model but also due to our model not fitting to all occurring cases. Lastly, we manually identify the differences which is a subjective process.

We use a model that describes the throw process regarding the strength of the acceleration. We expect two short and strong accelerations - first, an acceleration phase to accelerate the dart to an appropriate velocity and second a deceleration phase to stop the throwing arm again. We model this using a piecewise linear model with two subsequent peaks. In our experiment, we measured the accelerating forces at the wrist of the throwing arm according to three axes. We simplified this 3-dimensional time series data by computing the norm of the accelerating forces. Then we fitted for every throw our model to the time-series of the norm.

So, we fitted our model about the strength of acceleration to data about the strength of accelerating forces. One influence that occurs as accelerating force but which we did not model is the gravity. In Section 3.3, we already observed the effect of gravity to lift the baseline of the accelerating forces in rest phases. We could compensate this for our amplitude features by using the difference to the baseline for characterizing, e.g., the height of the first or the second peak. Yet, there is another influence which appears to displace certain form features. In Figure 6.3, we see again a time-series of the norm of the accelerating forces for a throw as a black line and the fitted model as a red line. Along the horizontal axis, we see the time in milliseconds in an interval with a length of 2000 milliseconds. Along the vertical axis, we depicted the strength. Additionally, we can see the intermediate versions of the fitted model with dotted lines. According to the data as well as the model, we can see a baseline at a strength of about 10. We clearly see the two peaks for the acceleration phases. Yet, we also observe a stronger mismatch between model and data between the two peaks and even stronger after both peaks or in other words at the time 0. Here, we see that the data falls below the baseline. Our model cannot follow because it assumes the baseline to be the minimum. We suppose

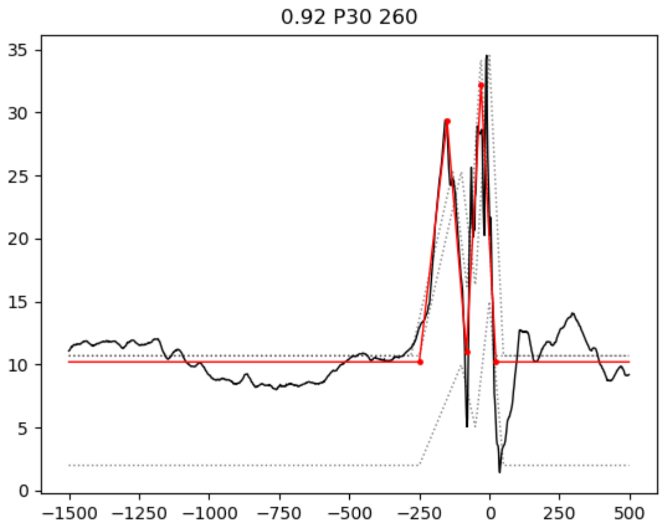

Figure 6.3: Annihilation of forces - Between both peaks and after the second peak at time 0 the norm of the measured forces fall far below the baseline which the model cannot represent.

that this is an effect of measuring accelerating forces from the throw process together with gravity. With our sensor, we measure the sum of accelerating forces over time according to three axes. One accelerating influence is gravity that is globally pointing downwards and static with a fixed strength. In phases in which the sensor rests or in phases of weak accelerations and slow velocity - e.g., when the participants slowly lift the arm to the initial throwing position - it is the strongest force that superimposes others. For this reason, we observe in those phases a baseline that is lifted to the static value for the strength of gravity. Yet, in phases of strong acceleration and deceleration, the related forces superimpose the effect of gravity. For this reason, we observe our two peaks. Yet, both forces are often not additive. For example, at the end of the phase of deceleration, some participants are moving their throwing hand almost vertically. They decelerate by carrying out a decelerating force that is directed upward. The sensor measures the sum of these force vectors. While gravity is pointing downwards, decelerating forces with a similar strength are pointing up. The summed vector that affects the sensor is shorter - or in other words, has a smaller strength - than gravity. In other words, those forces annihilate each other. For the norm of the accelerating forces at that moment it means that the norm has a value beneath the baseline which is what we can see in Figure 6.3. This annihilation depends on the directions of the occurring forces. Thus, it is specific to the throw pattern of participants. So it may occur for some participants and may not occur for others. The consequences are that the shape of the time-series contains unexpected features which our model cannot handle. For this reason, certain form features can deviate from their intended position and our model is less robust to be fitted to the data, which may, later on, influence the distribution properties by displacing median and increasing IQR. These form features are the basis for the throw features and influence therefore further analyses.

Another case in which the fitting of our model is less robust occurs when both peaks for the acceleration phases are badly separated. In Figure 6.4 and Figure 6.5 we have the data and the fitted models for two subsequent throws of the same participant. The course of curves for the data in both figures is similar with a small peak for the acceleration phase at time -100 ms, a big peak for the deceleration phase at time 0 ms, and an unclear transition between both. While both time-series are similar, the fitted models vary in this interval. In the following throw which we can see in Figure 6.6, the first peak could not be identified. Though our measures - median and IQR - are more robust against outliers, if the frequency of misfits is too high then this will affect median and IQR on these features presumably shifting median and increasing IQR.

Lastly, in our analyses, we manually identified the differences between good and bad groups according to the diagrams that we created. Yet, this is also a subjective and sequential process in which we may change our subjective criteria for identifying a difference. On time we may have accepted a visual pattern as a difference. Another time, we may have ignored a similar visual pattern. We see that there are multiple influences that may reduce the reliabliity of the results of our method. Yet, their extent is unknown so far.

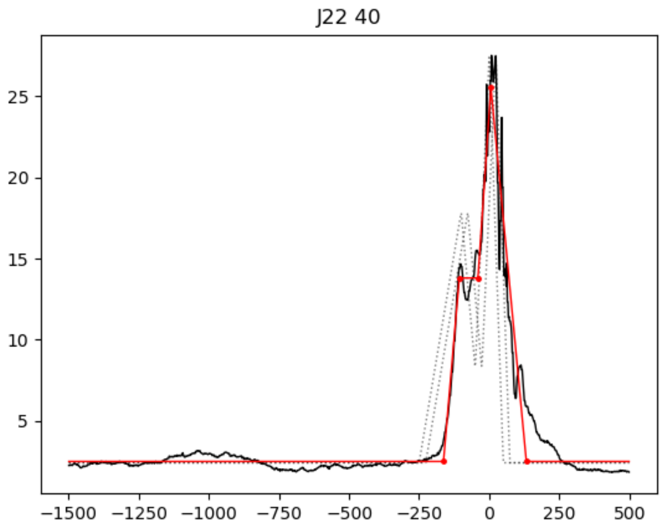

Figure 6.4: 40th throw - The model (red) could not be fitted to capture the valley between both peaks in the data (black).

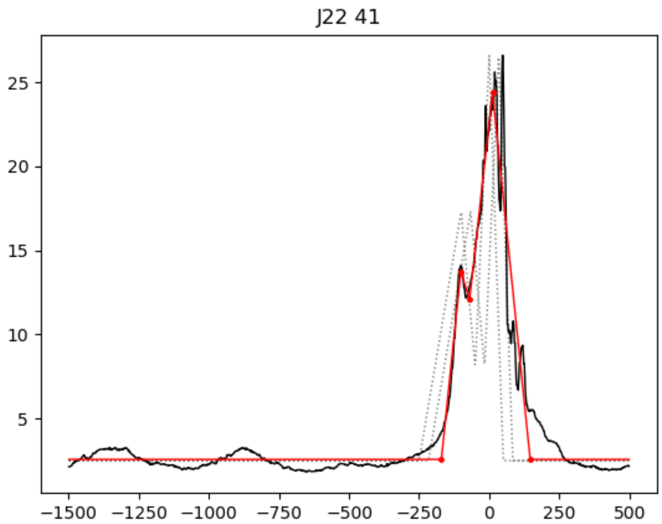

Figure 6.5: 41st throw - The model (red) could be fitted to capture the valley between both peaks in the data (black).

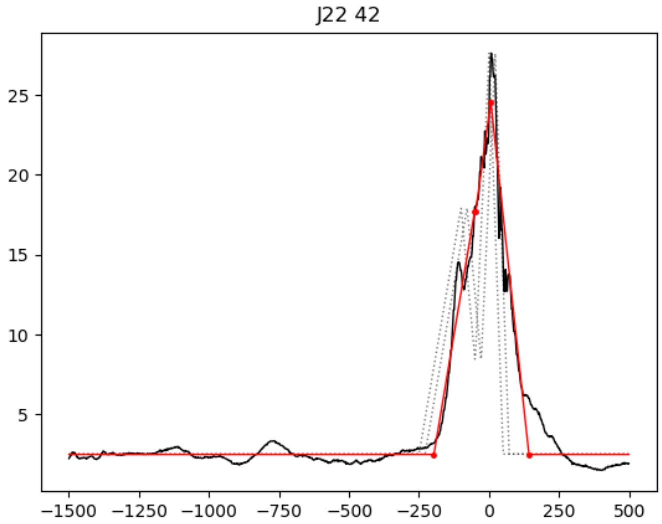

Figure 6.6: 42nd throw - The model (red) could not be fitted to capture the first peak for the acceleration phase and the valley towards the peak for the deceleration phase in the data (black).

Nevertheless, we managed to base our analyses on relevant data from a dart experiment in which we recorded data regarding performance and physiology of throws. Based on this data, we managed to make multichannel time-series of throws comparable by capturing kinematic features of throws. As we intended, we found and applied a way to quantify performance and physiology regarding darts-game, as well as the development in those measures. We also managed to apply multiple performance measures in a combined way in our analyses. Lastly, by assessing the relation between performance-related and physiological measures, we were able to identify the differences between high and low performing respectively strongly and weakly improving participants which we could compare against our expectations.

6.3 Future Work

In this last section, we like to connect this thesis to future research by providing ideas for future work for improving and extending our method, and to answer related questions. First, how may we improve our method to answer the questions apart from the examination and handling of the problems that we mentioned in Section 6.2?

To seperate the compound variances for the different types of variability in order to explore the quality variability, [Boenke2018] propose to focus solely on throws that fulfill the goal in order to avoid trivial variability and noise in order to keep the effect of the movement stable. We see another option to explore those correlations by exploring the development over time. We assume that trivial variability as well as noise variability decrease over time while quality variability increases over time. As participants practice, they find more options to fulfill the goal. Applying more different movements that reach the goal increase the variance in the movement that we observe. Regarding noise variability, participants may get better in avoiding noise. They may start to time their movements with breathing and the heartbeat. This will decrease the variance of noise variability over time towards a certain minimum. Regarding trivial variability, participants may get better over time to focus on the given goal. Also here, the variance in the movement will decrease over time to a certain minimum. Following the idea that trivial variability and noise variability decrease over time and only quality variability increases over time, we should see for participants that decrease in trivial and noise variability initially faster than they increase in quality variability a certain v-shape of the development. As the decrease in trivial and noise variability superimpose in the sum, the sum initially decreases as well. Over time the changes in both variabilities get smaller until the increase in variance for quality variability superimposes and the sum of variances start to increase. This is just an example development. In Figure 6.7 we depicted for every participant the development of the IQR of the feature d_t1_t2 over the 5 blocks of our experiment. The red lines mark participants who show in the fifth block a higher IQR than in the first Block. We see for several participants a decrease towards the second or third block. We also see for some participants an increase in towards the 4th and fifth block. This is a tempting approach which leaves the question of how these developments relate to the development in performance.

We decided in this thesis to exclude the data that we recorded in the second phase of the execution of our experiment from further analyses for several reasons like mismatching scales and artifacts from recording. We could try to solve these problems to extend the dataset. We may then use the additional data to validate the differences that we observed here.

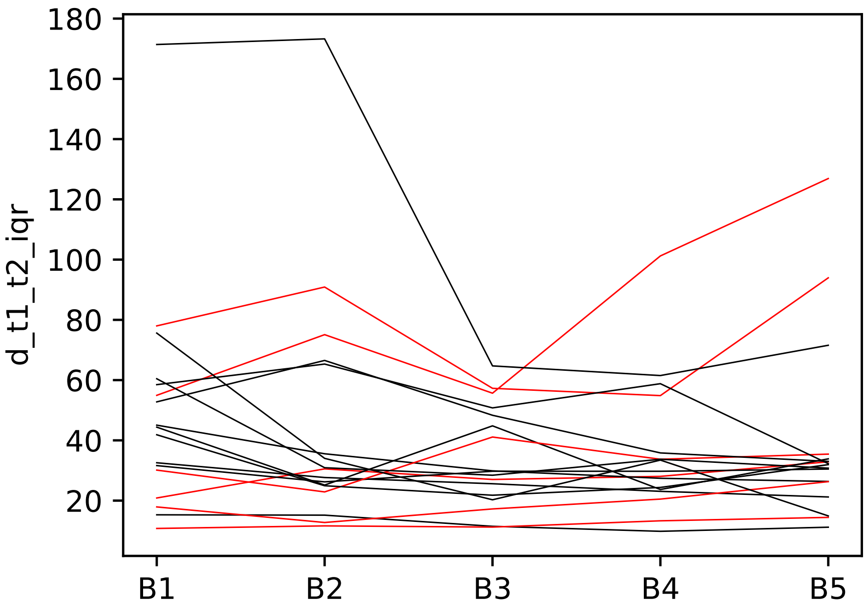

Figure 6.7: Development of IQR for d_t1_t2 over time

In order to reduce the errors that we introduce by abstracting the throw time-series data, we may search for a better model and improve the fitting process itself. Our model based on a piecewise linear function. Eventually, we could model the acceleration phases as bell curves or similar curves.

We could try to answer the research questions in different ways.

Following the idea of [Hulikanthe Math2017], we could explore the tendencies following a front from non-dominated sorting and compare the tendencies between good and bad fronts. This could provide us with hints regarding the influence of single performance criteria.

From the physiological data that we collected in our experiment, we focused solely on the strength of the accelerating forces. A different approach could be to explore in a similar way other measures like the rotational velocity. Next, we could also focus on multi-dimensional data. We could focus on the multi-dimensional kinematic data eventually defining a multidimensional model or reconstructing and analyzing the trajectories of throwing movements. Another multidimensional measure that we collected are the electrical activities from

EEG. With them, we could explore the relation between brain activity and performance or between brain activity and the variance in physiological throw measures.

In this thesis, we decided to apply differences to quantify development in performance measures and physiological measures. Though there is also the option of ratios as quantification for changes, we are curious about a multidimensional option. In this thesis, we already noticed the effect that is described by the law of practice. Depending on the progress of practice which is related to the performance the acquired improvements get smaller and smaller. To extract whether a participant improved much or few in comparison to others and in relation to the initial performance, we could determine the fronts from non-dominated sorting based on the performance and the development in performance.

In order to identify differences, we opposed good and bad groups which we selected based on the performance measures and the fronts from non-dominated sorting. We can think of other selection strategies. For development in performance measures, we have the concept of improvement and degradation. According to this, we could also oppose elements that improve with elements that degrade. So far, we followed the idea of selecting according to performance and observe the differences in the physiological measures. Another option is to select groups according to physiological measures and observe their differences according to performance measures. Summing up, future work can be done regarding different ways of quantifying physiology and performance, and other strategies to select groups for comparison to answer the same research questions.

From this work, we also met related questions.

Assuming that participants do not improve anymore, e.g., because they already reached some kind of personal maximum performance, we wonder if there is a tradeoff between our two performance measures. This is related to our next question.

In Section 5.1.1, we observed the triangle-like relation between our performance measures. We were able to show that our method does not evoke this pattern in general. So, what creates this pattern? Some of our considerations suspect some kind of feedback mechanism. With low precision, the hits of a participant are scattered all over the dartboard. This makes it harder to access a bias like our accuracy-error compared to the case of densely concentrated hits for a participant with high precision. In other words, it may be easier to perceive the extent of the accuracy-error and so to counter it when the precision-error is smaller. This may relate to the batch size that we use to aggregate our measures. We fixed our batch size to 100 which represents a block. Assuming that we would aggregate just 10 throws instead of 100, we suppose that we observe performance beneath the diagonal of the triangle that we observed. So, how does the batch size influence our results?

Since we considered the relation between performance and physiological measures, as well as the relation between the development of performance and physiological measures, a slightly different question would be the relation between initial status and development. So, how do good or bad performing participants change over time?

Lastly, We manually identified the differences between good and bad groups. We spend quite some time to do this. In addition, the results are affected by our subjective judgment. To improve this we wonder how we could automatize this process which may include the interfront and intrafront analyses.

Bibliography

[Ashby1958] W. R. Ashby. Requisite variety and its implications for the control of complex systems. *Cybernetica*, (1:2):83–99, 1958.

[Bernstein1967] N. Bernstein. *The co-ordination and regulation of movements*. Pergamon Press, Oxford, 1. engl. ed. edition, 1967.

[Boenke2018] Lars T. Boenke. Reframing variability: From nuisance to an aid to understand complex systems dynamics, Talk presented in the McIntosh Lab, Rotman Research Institute - Baycrest Centre, Toronto, Canada, 2018.

[Cheng *et al.*2015] Ming-Yang Cheng, Chiao-Ling Hung, Chung-Ju Huang, Yu-Kai Chang, Li-Chuan Lo, Cheng Shen, and Tsung-Min Hung. Expert-novice differences in smr activity during dart throwing. *Biological psychology*, 110:212–218, 2015.

[Deb2011] Kalyanmoy Deb. Multi-objective optimisation using evolutionary algorithms: An introduction. In *Multi-objective evolutionary optimisation for product design and manufacturing*, pages 3–34. Springer, London [u.a.], 2011.

[Edwards2011] William H. Edwards. *An introduction to motor learning and motor control*. Wadsworth Cengage Learning, Belmont CA, 1. ed., international ed. edition, 2011.

[Günnemann *et al.*2010] Stephan Günnemann, Ines Färber, Hardy Kremer, and Thomas Seidl. Coda. *Proceedings of the VLDB Endowment*, 3(1-2):1633–1636, 2010.

[Heylighen and Joslyn2001] Francis Heylighen and Cliff Joslyn. Cybernetics and second-order cybernetics. *Encyclopedia of physical science & technology*, 4:155–170, 2001.

[Hulikanthe Math2017] Nithish Hulikanthe Math. *Multi-Objective Success Factor Analysis for IT based Startups*. Master thesis, Otto-von-Guericke-University, Magdeburg, 2017.

[Latash2008] Mark L. Latash. *Synergy*. Oxford University Press, Oxford and New York, 2008.

[Lee *et al.*2014] S. Lee, Y. S. Kim, C. S. Park, K. G. Kim, Y. H. Lee, H. S. Gong, H. J. Lee, and G. H. Baek. Ct-based three-dimensional kinematic comparison of dart-throwing motion between wrists with malunited distal radius and contralateral normal wrists. *Clinical radiology*, 69(5):462–467, 2014.

[Levin2014] Andreas Levin. *Analyzing startup performance: How methods of machine learning and multi-criteria optimization may help to uncover successful startups*. Master thesis, Karlsruhe Institute of Technology, Karlsruhe, 2014.

[Preim *et al.*2009] Bernhard Preim, Steffen Oeltze, Matej Mlejnek, Eduard Gröeller, Anja Hennemuth, and Sarah Behrens. Survey of the visual exploration and analysis of perfusion data. *IEEE transactions on visualization and computer graphics*, 15(2):205–220, 2009.

[Ranganathan2004] Ananth Ranganathan. The levenberg-marquardt algorithm. *Tutorial on LM algorithm*, 11(1):101–110, 2004.

[Seidler *et al.*2017] Rachael D. Seidler, Brittany S. Gluskin, and Brian Greeley. Right prefrontal cortex transcranial direct current stimulation enhances multi-day savings in sensorimotor adaptation. *Journal of neurophysiology*, 117(1):429–435, 2017.

[van Beers *et al.*2013] Robert J. van Beers, Yor van der Meer, and Richard M. Veerman. What autocorrelation tells us about motor variability: insights from dart throwing. *PLoS ONE*, 8(5):e64332, 2013.

[Vrugt *et al.*2007] Jasper A. Vrugt, Jelmer van Belle, and Willem Bouten. Pareto front analysis of flight time and energy use in long-distance bird migration. *Journal of Avian Biology*, 38(4):432–442, 2007.

[Weber and Doppelmayr2016] E. Weber and M. Doppelmayr. Kinesthetic motor imagery training modulates frontal midline theta during imagination of a dart throw. *International journal of psychophysiology : official journal of the International Organization of Psychophysiology*, 110:137–145, 2016.

Declaration of Authorship

I hereby declare that this thesis was created by me and me alone using only the stated sources and tools.

Thomas Hennig

Magdeburg, November 20, 2018