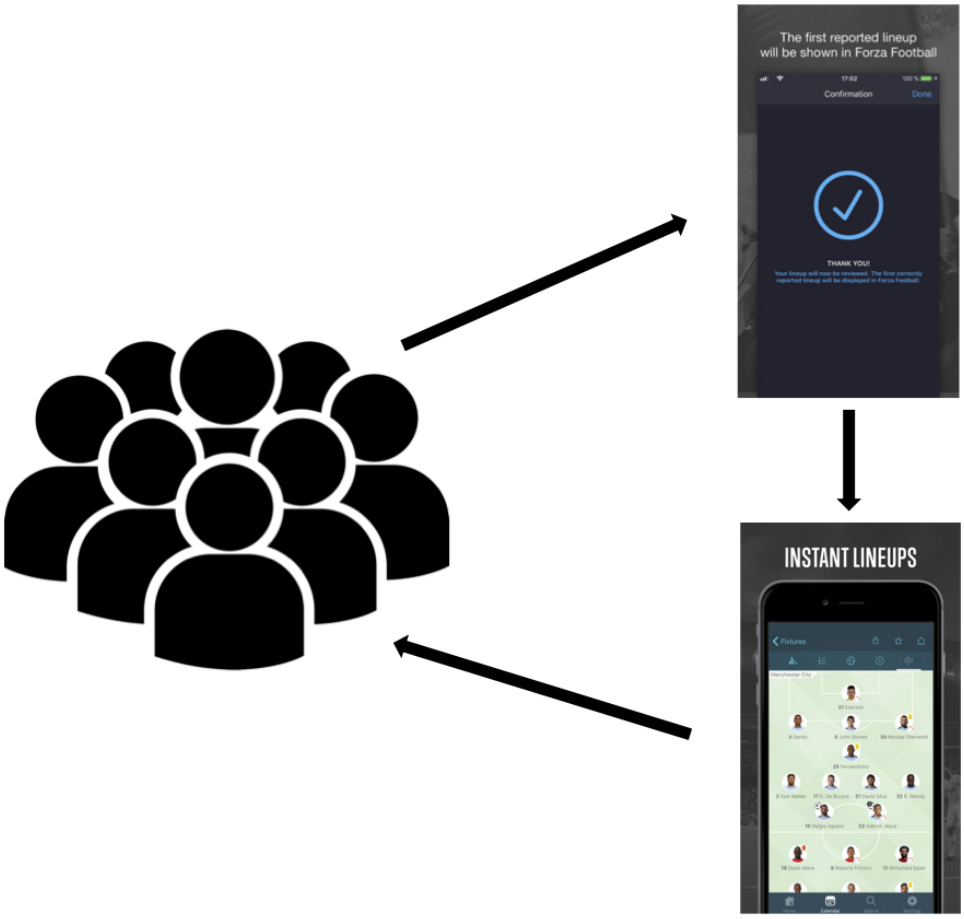
CHALMERS
UNIVERSITY OF TECHNOLOGY

UNIVERSITY OF GOTHENBURG



A study of how trust can be assigned in crowdsourced information retrieval

Master's thesis in Software Engineering

EMINA HROMIC

Department of Computer Science and Engineering
Chalmers University of Technology
University of Gothenburg
Gothenburg, Sweden 2018

Master's thesis 2018

A study of how trust can be assigned in crowdsourced information retrieval

EMINA HROMIC

Department of Computer Science and Engineering
Chalmers University of Technology
University of Gothenburg
Gothenburg, Sweden 2018

A study of how trust can be assigned in crowdsourced information retrieval
EMINA HROMIC

Supervisor: Jan-Philipp Steghöfer, Computer Science and Engineering.
Examiner: Regina Hebig, Computer Science and Engineering.
Advisor: Erik Heinemark, Forza Football AB

Master's Thesis 2018
Department of Computer Science and Engineering
Chalmers University of Technology and University of Gothenburg
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Gothenburg, Sweden 2018

A study of how trust can be assigned in crowdsourced information retrieval
EMINA HROMIC
Department of Computer Science and Engineering
Chalmers University of Technology and University of Gothenburg

Abstract

For establishing the trustworthiness of information retrieved through crowdsourcing, trust management systems can be used, which compute a trust value used in automated decision making. Using such systems, makes feedback vital since it gives users incentives to do more and better [1]. An instance of the more general problem is experienced by Forza Football, which has a desire to make automated decisions regarding users reported lineups. The purpose of this study is to explore possible ways to determine whether or not information retrieved through crowdsourcing from external entities is trustworthy.

For the purpose of this study, the Design Science Research methodology was used, beginning with a literature review for creating awareness of the problem along with gathering requirements from the crowdsourced information retrieval domain, followed by designing the trust management system, conducted through four iterations by evaluating it through validation sessions with relevant participants.

The results in this study indicate that using the trust value, to evaluate the trustworthiness of the information retrieved, does not yield similar results as a manual evaluation. When quality is of greatest importance, the data should be compared, for evaluating if several users have provided the same information. In those cases, a threshold can be set for which users' trust values needs to exceed in order to get their data approved. For lower quality controls, the trust value alone can be used for automated decisions. The feedback should be immediate, containing reasons for why the data was accepted or rejected, not including the trust value.

Keywords: Trust, Trust Management System, Crowdsourcing, Feedback.

Acknowledgements

This thesis was conducted contemporaneously with an employment at Forza Football AB, and was hence a collaboration with the organization. I would like to thank all co-workers for valuable brainstorming sessions and for giving me valuable wisdom and guidance in my work. I would furthermore like to thank all participants taking part in the various validation sessions, providing both good data and feedback. Additionally, I would like to thank the interviewees for providing valuable feedback and sharing their knowledge.

Last but not least, I would like to thank my supervisor at Chalmers, Jan-Philipp Steghöfer, for providing both great guidance and encouragement throughout the process.

Emina Hromic, Gothenburg, June 2018

Contents

List of Figures

List of Tables

1
Introduction

Information is a central part of human life, and the sharing of information and resources has become more important than ever, in particular in information systems and applications. The trustworthiness of the information provided in such systems is crucial, both for the system to operate in the right manner, but also for the social acceptance of it. Crowdsourcing is a typical concept of sharing information, which Jeff Howe defined as the act of an organization who outsources a function to entities external to the organization [5]. Crowdsourced information retrieval is a concept in which organizations take advantage of other entities by opening up their system giving the entities the possibility of providing information to the organization [5], [6]. This openness, allowing entities to provide information at any time, can cause great risks. It could furthermore provide a negative effect on the quality of the system's computation, if the information provided is not monitored in some sense [7]. One way of preventing this, which is of great importance, is to establish trust between the organization and the entities providing the information. This is due to the fact that this trust will increase both the robustness and efficiency of the system by having the uncertainty received by exposing the system to external entities, being decreased [8]. The concept of trust is the basis of many relationships and cooperation between parties, and is thus of great importance [9]. Trust can have numerous meanings, which can depend on the context or even the feeling of security [9]. One example trust is the one given by Diego Gambetta [10]:

"*Trust is a particular level of the subjective probability with which an agent assesses that another agent or group of agents will perform a particular action, both before he can monitor such action (or independently of his capacity ever to be able to monitor it) and in a context in which it affects his own.*"

Another definition of trust is provided by Jøsang, in his article, Can We Manage Trust, which is the following: "*Trust is the extent to which a given party is willing to depend on something or somebody in a given situation with a feeling of relative security, even though negative consequences are possible.*" [11].

Both these given definitions are based on an actor A having positive believes about actor B, who actor A depends on for its own welfare. It is furthermore evident that trust is an asymmetric relationship, and therefore, trust management can be seen from a two-sided perspective. Establishing trust between remote entities and assessing the trustworthiness of the information provided by a remote entity is a difficult task, since it cannot be done in the same manner as the traditional way,

when entities actually can meet each other.

In order to deal with the issue of trust when using crowdsourcing for information retrieval, to establish the trustworthiness of the information retrieved from the entities, trust management systems can be used. These kinds of systems make automated decisions based on a certain value of trust that each entity has been given based on previous performances. The difficulty is how this trust value should be assigned to the entities and how it should be used by the trust management system. Thus, the purpose of this study will be to find a way to assign and use trust in crowdsourced information retrieval. A trust management system will be developed as a means for the study.

This study will be a collaboration with Forza Football AB, which has a requirement to use automated decision making in their application Forza Reporter. Forza Reporter is a tool used for allowing users to report lineups for numerous football teams, for games being played on the same day. The solution for tackling the requirement is to integrate a trust management system, which assigns trust values to users based on previous performances and automates the decision of whether or not the user's reported lineup should be approved based on that trust value.

The study aims to contribute by exploring possible ways to determine whether or not information retrieved through crowdsourcing from entities external to an organization is trustworthy. One of the issues here is to decide when someone actually can be seen as trusted and how a trust value should be assigned given certain circumstances. This is particularly evident when trust is intended to be used as the basis for automated decision making. Therefore, the study furthermore aims at exploring how a trust value should be computed and then used. With the use of trust values, it is vital to have some kind of feedback, letting users know why a certain trust value is given to them, what the value means, and what is required by the user in order to reach a certain level. Feedback is of great importance, because it shows users that the information they provide actually is being used, or at least acknowledged. Furthermore, feedback is a key for giving users an incentive to do more and better [1]. Thus, another contribution from this study will be the exploration of how and what kind of feedback users should receive regarding what the given trust value mans and what is required by the user in order in order to improve the trust value.

1.1 Project goal

The purpose of this study is to explore possible ways to determine whether or not information retrieved through crowdsourcing from entities external to an organization is trustworthy. This will be the basis for further exploration of how the trust can be assigned and then used for making automated decisions. To enable this, a trust management system will be developed. The system will be developed for the Forza Reporter application, with the purpose of assigning trust values to users based on previous performances, in order to distinguish the trustworthy users from those

who are not. The organization Forza Football AB has a desire to in the future use the results from this study in order to automate the decision of whether or not their users' reported lineups should be approved.

This study aims in particular at answering the following research questions:

- **RQ1:** *Does using the trust value of a user to evaluate the trustworthiness of the information provided from that user yield similar results as a manual evaluation?*
- **RQ2:** *How should the trust value be used once assigned to the user?*
- **RQ3:** *How should the user receive feedback regarding what should be done to reach a certain trust value?*

1.2 Scientific contribution

This study aims at providing other practitioners, such as application developers adapting crowdsourcing, guidance in what various ways users can be distinguished using trust, in particular by integrating trust management systems into their current systems. The research existing at this point brings up quite extensive information about crowdsourcing in general, including its advantages and disadvantages. There is also research on trust management systems, where some authors provide concrete algorithms for how the trust management systems suggested by them can be used. What however is lacking in today's research, is how trust management systems can be adapted in crowdsourcing. This study thus aims at providing ways of how trust management systems can be integrated into crowdsourcing applications in order to distinguish trustworthy users from those who are not.

Along with this, the study furthermore aims at providing these practitioners with proper ways of providing feedback to their peers, in particular for those being in the same situation as Forza Football. This since larger parts of existing research provides guidance for how to give feedback to students, co-workers etc., but not in particular how to provide feedback to users taking part in a crowdsourcing application, that uses a trust management system, meaning that, should the trust value be displayed to users or is the value insignificant to them? Or should the value be gamified in some manner, adapting it more to the users and making it more understandable to them? Or should the feedback simple not include this at all? These are examples of questions which existing literature does not answer. Therefore, the results from the evaluations taking part in this study will give propositions of what kind of feedback is good feedback for users in such a context.

1.3 Research Methodology

For the purpose of this study, the Design Science Research methodology is used, as suggested by Hevner [12] as shown in Figure 1.1. The methodology consists of

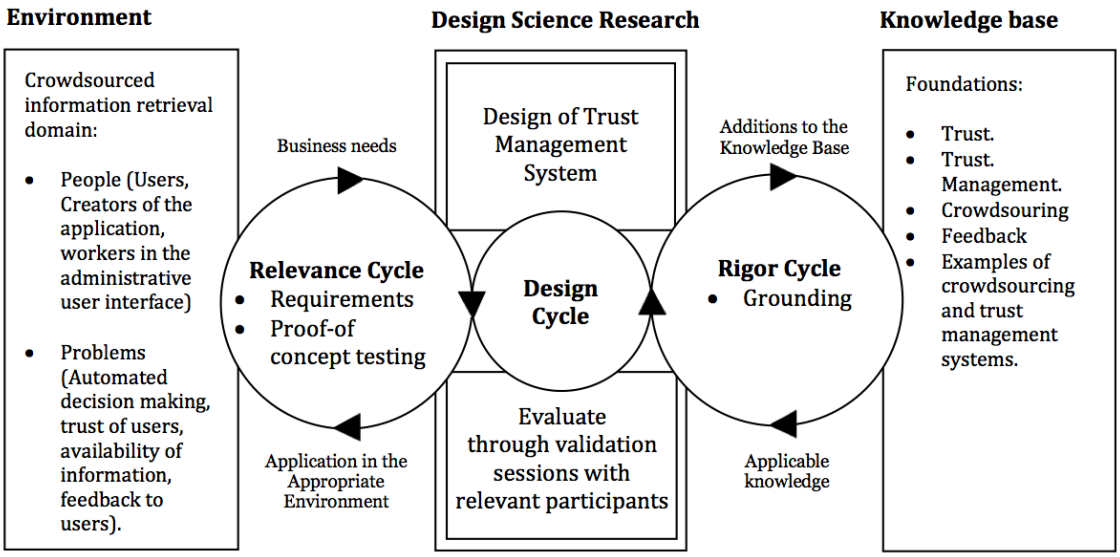


Figure 1.1: Design Science Research cycles for this study. Model adapted from [2]

three main components, combined by three different cycles, which are to be completed iteratively. The components are: the knowledge base, the environment and the actual design science research. The three connecting cycles are: the rigor cycle, the design cycle and the relevance cycle. It furthermore mainly addresses research by development and evaluation of IT artifacts that are designed to meet a certain need, particularly within the industry, just as in the case of the organization Forza Football.

The rigor cycle is the connector between the knowledge base of the foundation and methodologies with the activities occurring within the design science [2]. During the initial steps of the methodology, an extensive literature review was conducted in order to create awareness of the problems existing, but also to provide a foundation for the contribution of this study. The knowledge base applicable for this study consists of: Trust, Trust Management System methodologies, Crowdsourced Information Retrieval, Trust Management Systems in Crowdsourcing, the Process of Feedback and Examples of Crowdsourced Information Retrieval Systems.

The relevance cycle is the binding component between the environment of the research project with the activities taking place in the design science research [2]. Within the Crowdsourced Information Retrieval domain, the environment consists of the creators of the application and the users of it. Furthermore, it includes the existing problems such as the lack of automated decision making, trust in users (both assigning trust, but also retaining it), feedback to users when reporting a lineup and the availability of information.

Within the design cycle, the knowledge gathered prior to this state is applied in order to design and develop the trust management system, to make sure that this can be applied and solve the problems within crowdsourced information retrieval. In order to validate this during the design and development of the trust manage-

ment system, it will be evaluated in the various iterations in the design cycle, by conducting validation sessions with entities relevant for this study.

How this research methodology has been used during this study, in order to answer the research questions previously presented, will be presented further in Chapter 8.

1.4 Limitations

The research on trust can be put into three different categories: microanalyses, mesoanalyses and macroanalyses [13]. The former refers to the interactive generation of trust from the individual standpoint. The second category refers to research on trust in social environments such as teams, partnerships, organizations etc. And the latter case refers to research on the impact of trust in social systems and the entire society [13]. Since this study concerns trust between an organization and the entities outside the organization, the research will be limited to the research of mesoanalyses.

With the current version of Forza Reporter, there does not exist any possibility of providing feedback to users, except writing to them directly, and manually, through an integrated messenger inside the application. In order to provide users with feedback about their reported lineups, and provide them a list of their historical reports, which is a form of feedback itself, it is essential to have user profiles. Even though users are unique by their user ID internally, user profiles are vital and required for the process of feedback to be possible. This due to that the user ID is only a unique distinguisher, used internally to keep track of users, and does thus not provide any kind of way of giving users feedback. Since user profiles do not exist in the version of Forza Reporter prior to the beginning of this study, this becomes a limitation. Due to internal limitations at Forza Football AB, it is not possible to integrate user accounts into Forza Reporter during the time limit of this study. Therefore, a workaround will be have to be made. Different drafts will be constructed, with different options of potential feedback, which will be evaluated together with various participants in order to conclude what type of feedback is the best one for this particular case.

1.5 Report structure

The report has been structured in such a way that it first provides the problems that this thesis aims at solving. This is followed by extensive background information relevant for the topic of this thesis. After giving the reader enough background knowledge, the design of the trust management system is presented, followed by the results, a discussion and conclusions.

The initial part of the report provides the problem statement. The reason for providing the problem statement so early in the report, is due to that the extensive

background information provided afterwards, is provided in the context of the problem being solved in this study, making this order natural. Within this chapter, existing problems with trust and trust management in general will be presented, laying a foundation for the more concrete problem for which this study revolves around. This study aims at solving Forza Football's problem with trust towards their users, knowing how to distinguish trustworthy users from those who are not, and using this as a basis for automated decision making. With that comes a trust management system. The trust management system is to be used, in conjunction with the current system, in order to assign trust values to users depending on their previous performances, and thus distinguish the users with positive outcomes from those with negative. More specific details about the existing problems will be presented within this chapter. Furthermore, in conjunction to this, the requirements for the trust management system for this study will be provided, in order to provide information about how the trust management system for this study is intended to be used.

Following the problem statement comes extensive background information chapters, which has been divided into five different chapters, relevant for each topic. The third chapter, *What trust is and how to manage it using trust management systems*, provides an insight into the concept of trust and the definitions if it. The chapter furthermore provides a definition of what trust management is, how trust can be managed, and different kinds of specializations of systems adapted for trust management. Thereafter follows a section about which specialization of the different trust management systems has been chosen for this study, and the reason behind this choice. The fourth chapter, *Crowdsourced Information Retrieval*, provides definitions of crowdsourcing, examples of concepts commonly mistaken for crowdsourcing, different categorizations of crowdsourcing cases within different kinds of typologies and issues existing within crowdsourcing. The fifth chapter, *Trust Management Systems in Crowdsourcing*, presents the opportunities of trust management systems in crowdsourcing. In the sixth chapter, *What Is Feedback And Why Is It Important*, the different strategies and content, important for good feedback, are a big part of this chapter. The chapter is concluded with reasons for why feedback is important. The final chapter in the literature review, *Examples of Crowdsourcing Information Retrieval Systems*, provides examples of Crowdsourced Information Retrieval Systems, both those that are completely crowdsourced, but also those applications adapting crowdsourcing by integrating trust management systems into their current systems.

Within the eighth chapter, the design of the trust management will be presented, followed by the results retrieved from validation sessions conducted with participants relevant for this study, ending with a discussion and conclusions.

2
Problem Statement

In order to deal with the issue of trust when using crowdsourcing for information retrieval, to establish the trustworthiness of the information retrieved from the entities and the entities themselves, trust management systems can be used. When using trust management systems for automated decision making, users are given a certain trust value, depending on their previous performances. One of the difficulties here lies in knowing when a user can be seen as trusted, and thus knowing how the trust value should be assigned to the user and how it should be used by the trust management system. If a user has a high trust value, all the information within the context in question, provided by that specific user will automatically be trusted. Whilst information coming from a user with a low trust value will require to be reviewed cautiously before an automated decision can be made. Another issue existing when using external entities for information retrieval is the veracity and validity of the information. The veracity of the information refers to the quality of it, which refers to the level of uncertainty of the information retrieved due to the information being inconsistent or incomplete. The veracity of the information retrieved is derived from the user's trust. The validity of the information retrieved refers to how correct and accurate the information is for the intended purpose [14], [15].

Trust is a relatively sensitive concept, and yet it is one of the most important aspects in human life, laying the basis for numerous kinds of relationships [5], [6]. Numerous researchers have throughout the years tried defining the concept of trust, especially in different kinds of contexts, and yet they have only managed on agreeing on certain common characteristics that trust possesses, such as that it is based on experiences with other peers and its relevance when facing risks. The most common characteristics defining trust are the following:

- **Trust is implicit**: For a trust relationship between the trustor A and the trustee B, it is possible to define both the context and the timespan related to this relationship. It is however not possible to explicitly define the willingness and capability of other entities potentially involved in this relationship. It is also not possible to estimate changes in the context, and as the timespan changes, so does ones beliefs [16].
- **Trust is subjective**: If trustor A trusts the trustee B, it does not necessarily mean that entity C also will trust B. Trust therefore depends on the trustor's perspective.
- **Trust is asymmetric**: If trustor A trusts the trustee B, it does not necessarily mean that B will trust A in turn. This means that that trust is a non-mutual

reciprocal. Despite this, trust can at times be symmetric.

- **Trust is context-dependent**: Having trustor A and trustee B, the trust between these two entities may depend on different contexts, such as task goal, timespan or environment. This means that trustor A may trust the trustee B for a particular task goal, within a given timespan in environment X, but not in environment Y, for example.
- **Trust is antonymous**: Trust depends on the given context. This context may however be perceived differently by the trustor A and the trustee B [16].
- **Trust may not be transitive, but propagative**: If trustor A trusts the trustee B, and B in turn trusts another entity C, it does not necessarily mean that A also will trust entity C. However, the trusting relationship between entity B and C, can give entity A an indication of the trustworthiness of entity C, in order to make an assessment of the trust of entity C.

These characteristics add a complexity to trust in which many aspects need to be taken into consideration when deciding whether or not to trust another entity. Some of the issues existing are deciding when someone actually can be seen as trusted and how this trust can be managed. This is particularly evident when trust is intended to be used as the basis for automated decision making. One of the greatest problems here is what defines an entity as trusted. Another problem is what possible ways there are to determine whether or not information retrieved through crowdsourcing from entities external to an organization is trustworthy. This especially since trust can vary depending on the context, if it is trust between humans or a human and a system or social aspects, just to mention a few [9]. For this study, it is of particular interest what defines a user as trusted and how that trust can be established. Furthermore, in trust management, each user is usually given a certain trust value, which is a value indicating users trustworthiness based on previous performances. The problem with such a value is how it should be computed and used, once assigned.

2.1 The Forza Reporter problem

An instance of the more general problem of automated decision making based on crowdsourced information, is the one that Forza Football AB is experiencing with their application, Forza Reporter. The application, established in August 2017, is a football application relying on crowdsourced data. It allows its users to report lineups for all teams existing in the application. If the lineup is approved, it will be displayed in their live-score application, Forza Football, containing statistics about both teams and players [17]. The process of approving a lineup in today's situation is not optimal and relatively time consuming. The solution used currently is an administrative user interface tool, where all reported lineups are received. In the administrative user interface, depending on if the reported lineup is with or without formation, there is an overview of the starting eleven on the field or a list representation showing the starting eleven together with the substitutes, respectively. Within the administrative user interface, a person needs to examine each reported lineup by comparing the user's reported lineup with the source that the user has provided.

This source is required to be official and should indicate from where the user found the lineup. A valid source is for example the club itself, well-known newspapers or similar. This means that, within the administrative user interface tool, the source that the user has added needs to be reviewed as well upon receiving the lineup. If the lineups are matching and the source is valid, the lineup will be approved and thus displayed within the application Forza Football. If the source is invalid however, the lineup is immediately discarded. Whilst in the case of any incorrectness within the reported lineup, the decision of approving or discarding it will depend the severity of the incorrectness. If there are a few players that are incorrect, which may be due to that some players are missing in the squad in the application, or if few players' positions have been switched, the lineup will be updated to match the official one by the persons working in the administrative user interface and then approved. A few players means perhaps 2-3 maximum. On the other hand, if half of the team or more is incorrect, the lineup will be discarded immediately. Furthermore, with today's solution, the users do not receive any kind of feedback regarding whether or not their lineup was approved or any reason for why their lineup has been rejected.

During the few months that the application has existed, numerous lineups have been sent in by users. The data collected from this has indicated that several, typically different, users either have reported the completely wrong lineups or had invalid sources, for the same game. The conclusions made so far are that there currently is no proper incentive for users to report valid lineups and no methodology for Forza Reporter to assign trust to users and therefore distinguish those who have a positive impact, meaning those users who report correct lineups, from those who do not. Furthermore, there is a great need for automated decisions completed by a system, in Forza Reporter. This due to that there currently are many users and lineups streaming in, into the administrative user interface, where workers at Forza Football AB need to spend numerous hours reviewing lineups and making a decision for what to do with them. The application is currently only released in three countries, but is planned for being released in several more, which immediately means more users and more lineups, which would become work overload for the team responsible for the received lineups. In order to solve these existing issues, there is a need for a trust management system for Forza Reporter. The trust management system is to be used, in conjunction with the current system, in order to assign trust values to users depending on their performances, and thus distinguish the users with positive outcomes from those with negative ones. The desire from the organization Forza Football AB is to use this trust management system in order to make automated decisions regarding users reported lineups. The aim is to have the trust management system determine users trust values based on previous performances and make a decision based on the value: either to approve the lineup reported by the user, or to discard it. The trust management system should also increase or decrease the users trust value, depending on if the user reported an accurate lineup or a completely inaccurate one, respectively.

2.2 Requirements for the trust management system for this study

The trust management system that is to be implemented for the purpose of this study, needs to meet certain requirements. The trust management system is to be used, in conjunction with the current system, Forza Reporter, in order to assign trust values to users depending on their performances, and thus distinguish those users with positive outcomes, meaning those who provide accurate lineups, from those with negative ones, corresponding to users who provide inaccurate lineups. Users' performances are decided by the historical data available of the lineups they have reported and the quality of each reported lineup. As mentioned in Section 1.1, the organization Forza Football AB has a desire to in the future use the results from this study in order to automate the decision of the approval or rejection of users' lineups. It is therefore a part of this study to investigate how the trust management system will be used in order to make automated decisions regarding users reported lineups, based on the trust value. The trust management system should also increase or decrease the users trust value, depending on if the user reported a correct lineup or a completely incorrect one, respectively. The increment of the trust value should furthermore depend on the quality of the lineup reported by the user. Once the trust value has been assigned, the problem that arises is how the trust value should be used and what metrics, besides the outcome of the reported lineup, causes the trust value to be increased or decreased. The most important aspects needed to be taken into consideration when assigning the trust value to a user and then putting it into use, are the following:

Assigning trust values

- Assign trust values to users based on previous performances. The value lies between 0-1. 0 indicates that all the user's reported lineups have been rejected. Having 1 in trust value indicates that the user's Z most recent lineups have been accurate, meaning that the user is seen as completely trusted. How many lineups, Z, need to be reported will be tried out during the different iterations in this study.
- When reporting the first correct lineup, what trust value should the user receive?
- What happens when the user reports an incorrect lineup?
- Should the user report x correct lineups in order to get a higher trust value? Or does the trust value increase with a certain value X for each correctly reported lineup?
- Should the increase of the users trust value also depend on the correctness of the lineup? There are cases where the lineup provided by a user is approved, after certain, minor adjustments have been made. That would mean that the increase of the users trust value would thus depend on the number of adjustments required prior to the approval of a lineup. Or should the trust value be increased with the same value no matter the case?
- Increase the users trust value if the reported lineup is correct, even though the

lineup might not be approved.
- Decrease the users trust value with Y when an incorrect lineup is reported.
- Should Y depend on the severity of the error? Meaning, if the error is a matter of one incorrect position or player, the decrease of the trust value should be minor, and for the opposite case, it could be a matter of purposeful damage, and the trust value should then be decreased with a larger value.
- Should the trust value be global or local? Meaning, should the trustworthiness of a user be valid throughout the entire content in the application, or should it only be valid for the specific team for which the user has reported lineups for?
- Should the trust value for each user, be calculated based on a specific timespan, e.g. last two months, or from the users initial lineup until today?

Using trust values

- Should the system only rely on the user's trust value or by comparing users lineups for the same team with each other, when making automated decisions? Or both? Or does the automated decision making process need to be supported by another value, indicating the user's history of reported lineups in some manner, for example?
- Should a user get its lineup automatically approved first when its trust value is above a certain given threshold?
- If setting a threshold for what trust value is required for getting a lineup approved, what happens with lineups reported by users, if their trust values are below the threshold, but their lineups have proven to be correct after comparisons between each other?
- Should a user with a trust value beneath a certain threshold automatically get its lineup discarded? And leave room for it to be manually reviewed? Or should the user's lineup automatically be compared to other users lineup for the same team, and let the system make a decision based on their similarities?
- If comparing users lineups with each other in order to decide whether or not the lineup is correct, should the users' trust values decide who gets its lineup displayed in Forza Football, or should that be based on which user reported the lineup first?

Once this trust value has been put into use, the user should receive some kind of feedback regarding the reported lineups and trust value. Feedback is of great importance, because it shows users that the information they provide actually is being used, or at least acknowledged. Furthermore, feedback is a key for giving users an incentive to do more and better [1]. Thus, the feedback needs to be informative enough to give the users an incentive to continue reporting correct lineups. The feedback could include both a response of whether or not the user's lineup was approved, and depending on the outcome, what trust value the user is given. What needs to be considered with the feedback is the following:

Giving feedback

- Should the trust value be displayed to the user? Will the user understand what that number means? Or is this not significant to the user?
- Should the trust value be gamified? Meaning, if a user for example has a trust value of 0.1, the user is a water-boy. Or if the user has 1.0 in trust value, the user is a coach.
- Should the trust value be displayed together with its gamification to the user? Or only the gamified value?
- Should the user receive feedback about what is required to reach the next step or even the highest value?
- When the lineup is correct or the user's trust value is high enough to get the lineup automatically approved, send proper feedback explaining the reasons for the approval.
- When the lineup is incorrect, but the user has a high trust value, send feedback explaining the reason for rejection.
- When the lineup is incorrect and the user's trust value is low, send feedback explaining the reason for rejection.
- For each rejected lineup, should the feedback include the new trust value, the reason for it being lowered, and an explanation for what the user has to do in order to reach a higher trust value again, or only the reason for rejection?
- For each approved lineup, should the feedback include the new trust value and an explanation about what that new value means, or only the reason for approval?
- What proper feedback is, will be investigated during the different iterations in this study.

As it can be seen, the process of feedback is a complex one, requiring a lot of aspects to be taken into consideration. But at the same time, feedback is one of the most important things, and is thus also of great significance for this study.

3

What Trust Is and How to Manage it Using Trust Management Systems

Trust is the basis of many relationships and cooperation between parties, and is thus of great importance [9]. Trust can have numerous meanings, which can depend on the context or even the feeling of security [9]. During our lifetime, we will meet numerous people in different kinds of situations, giving us the opportunity to achieve our goals in life. The success of these different opportunities sometimes solely depends on our own efforts and are therefore dependent on our own responsibility. In other cases however, the opportunities we have may depend on other people's will and good-faith to complete what we expect them to. With interactions such as these, risks arise, since we need to completely rely on an uncontrollable situation, without any certainty of the final outcome [9], [18], [19]. In some scenarios, we make the decision to rely on someone else's actions, despite knowing the potential risks. In situations such as these, when the person we are relying on does not perform actions meeting our expectations, we may reach two different states. We are aware of the risks beforehand, but nevertheless decide to rely on someone else's decisions and actions. In the other case, we may come to feel disappointed and potentially even mislead, because we left our future in the hands of another person's decisions and actions, without concern of the potential risks ahead. This is the case that Forza Reporter currently is facing, relying on their users to provide them with lineups for teams, and thus facing both the risk of incorrect information and users who deliberately want to destroy for the organization behind the application, Forza Football, and reaching a point of disappointment due to that the outcome of the crowdsourcing project has not become as expected.

These different scenarios provide an indication of the value and importance of trust. Despite the ease of trusting another person, risks will always be involved. Even though risks are involved in the concept of trust, there are numerous other difficulties at hand. *What is trust? When can someone be seen as trusted? What does trust depend on? And what are the greatest issues existing with trust?* Numerous researchers have throughout the years tried to define the concept of trust, especially in different kinds of contexts, and yet they have only managed on agreeing on certain common characteristics that trust possesses. The research on trust can be put into three different categories: microanalyses, mesoanalyses and macroanalyses [13]. The former refers to the interactive generation of trust from the individual stand-

point. The second category refers to research on trust in social environments such as teams, partnerships, organizations etc. And the latter case refers to research on the impact of trust in social systems and the entire society [13]. Since this study concerns trust between an organization and the entities external to the organization, it is the mesoanalyses that is of interest.

3.1 The concept of trust

Trust is a concept that can take various forms, depending on the context, the feeling of security, the persons involved, just to mention a few variables that may affect our trust in someone else [11]. That we say that we trust our doctor to heal us, will not necessarily mean that we trust other drivers when being in traffic, and vice versa. And our trust in a user who provides correct lineups, does not mean that we will trust the same user in providing players in a squad, and vice versa. Saying that we trust a particular person in a particular situation, does not mean that we would trust the same person within another situation in a different context, or that we would trust another person in the same context. This means, that trust may have a different meaning depending on the persons involved and the context.

Trust requires at least two entities, and is thus normally seen from a two-sided perspective. An entity A relying on another entity B assesses the trustworthiness of B, while B, in some cases, is interested in giving the best picture of itself, by meeting A's expectations. In other cases however, there is a possibility of that B might intentionally or unintentionally try to exploit A. This can be achieved by giving A false impression of B's trustworthiness for whatever the reasons may be, such as for B's own personal gain [11]. Even though trust normally concerns the relation between two individuals, it may also extend further and thus concern the relation between an individual and other entities such as a group, organizations etc. The person trusting another entity, is normally referred to as a *trustor*, while the entity being trusted is referred to as a *trustee*. The trustor A normally only trusts the trustor B within a certain context X, which is referred to as the content of the trust relation. This relationship between the trustor and trustee is known as a *trust relation*. The fact that A (the trustor) trusts B (the trustee) with respect to X, is the basis of different trust relations, such as the trust relation between two partners, or the trust relation existing between an organization and its users taking part in a crowdsourcing project, providing information to the organization [3], [11], [20], [21]. For this study in particular, the organization Forza Football AB, is in a need of users that they can trust in the context of lineups. This trust relation between these two entities involves risks, both in the case of Forza Football AB, but also in general. When trustor A leaves its faith in the hands of the trustee B, there is no certainty that B will complete and deliver what is expected, and thus, A has no certainty of whether or not the expected outcome will be reached [9], [18], [19]. This is something that has been evident in Forza Reporter, where there has been all kinds of users who at some points meet the expectations, but at other times do not deliver what is expected. The consistent users, providing either correct or incorrect lineups at all times have created a pattern, letting the organization know what the

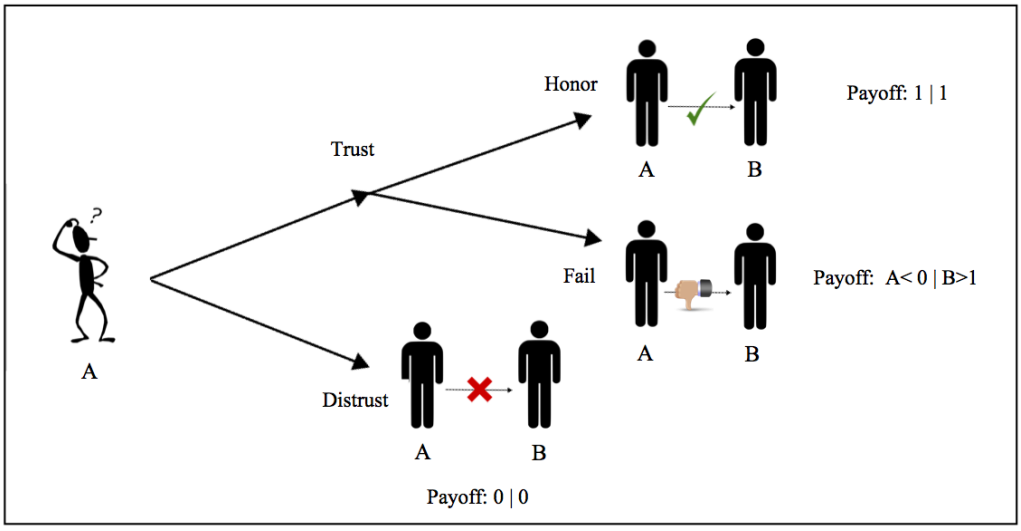


Figure 3.1: Model of the trust problem. Model adapted from [3]

outcome of the next lineup most likely will be. Whilst with the inconsistent users, their are great risks at hand, since there is no way for the organization to tell what they will provide next time they report a lineup.

In his book, *Trust and Rationality*, S. Alexander Rompf presents that the trustor faces a trust problem when facing a situation where a decision needs to made regarding whether or not another entity can be trusted [22]. In this trust problem, the trustor may make two different kinds of decisions, to either trust or distrust the trustee. Further work on the same topic was conducted by Wang and Singh, who build on top of Jøsang's definition of trust, adopting it into a tripled trust space, consisting of belief, disbelief and uncertainty. "*Trust in this sense is neutral as to the outcome and is reflected in the certainty (i.e., one minus the uncertainty)*" [23]. The first situation which may occur is **trust in a party**, where the belief in the trustee is high, the disbelief is low and the uncertainty of the trust relation is also low. Rompf defines the trust in a party to be an act where the trustor completely relies on the trustee, and may not predict with certainty how the outcome of the trustee's decision-making and actions thereafter will end up. The trustor can also not affect the outcome with external enforcement. The trustee's actions solely decide whether the trustor will loose or gain something from the trust relation. **Distrust in a party**, is the second situation, where the beliefs in the trustee are low while the disbelieves are high. The uncertainty in this particular situation is also low. Rompf adds to this, by stating that the trustor can affect the outcome of the interaction with the trustee with some degree of certainty, by eliminating potential damage from the start. In this case, the trustor does not solely rely on the trustee's decisions and actions as for determining whether the outcome will lead to a loss or gain. By having the trustor controlling the situation, the potential risks are significantly lowered, which gives the trustor a larger certainty of the expected outcome. The final situation existing according to Wang and Singh, is the **lack of trust in a party**, where the uncertainty of the trust relation is high. The degree of

uncertainty when trusting another entity, varies depending on two key metrics: 1. Effect of evidence: If the amount of trust evidence is high, so will the certainty be; and 2. If the conflicts between positive and negative outcomes is high in the trust evidence, the certainty decreases [23].

Rompf additionally states that, in most cases, the trustor has no other option than to trust, meaning that rarely, the trustor has the possibility of controlling the interaction completely, both in social and in computational systems. In the cases where a trustor A, completely needs to trust, trustee B, the actions completed by B would be called a trusting act, which is the constitution of the trust relation between the two entities. When there is a trust relation X between the two entities, it means that the trustee B has made actions that are of no harm for the trustor A, but rather have created gain for A. With such actions, B is seen to be trustworthy and thus fulfills the trust of A. In some cases, however, A has the possibility to affect the outcome with some degree of certainty, without solely trusting B and the actions completed. The relation between A and B would then be known as distrust. This trust problem, having a trustor A making the decision regarding if a trustee B can be trusted or not, with respect to X, is demonstrated in Figure 3.1. As mentioned in Chapter 2, Forza Reporter requires its users to add an official source of where the lineup was found, in order to make sure that the lineup is an accurate one, and not just a made up lineup created by the user. Forza Football AB currently also review all incoming lineups manually, which also is a way for the organization to make sure that no incorrect or misleading lineups are displayed in their live-score application Forza Football. This is a way for them to control the situation, as for what is displayed in Forza Football. There is however no way for the organization to control the actions completed by the users, meaning that they need to completely rely on the users, and trust them to provide what is expected.

Within the trust problem, there are different possible outcomes, as presented in Figure 3.1. One of the outcomes could be that the trustor A, distrusts the trustee B, which could lead to no interaction at all between A and B. This would give a payoff of zero for both parties since none of them would gain anything in this case. For the opposite case where the trustor A trusts the trustee B, it may lead to two different outcomes: honor and fail. When the former case is reached, it means that B has satisfied A's expectations leading to an equal payoff for both A and B - a win-win situation. What B would gain from such an outcome, varies depending on the situation and trust relation, but it could be anything from recognition to important contacts, money, internship, just to mention a few. For the other case however, if the trust relation leads to failure, it means that B has failed A and thus violated the trust relation. The payoff for A becomes negative, since A has put its trust into B, where B has failed to meet the expectations. In this situation, the trustor A has put more on stake than B, and thus looses more than B does, whilst B still gains something, leading to a payoff smaller than one.

Example: In the case of Forza Reporter, where the organization Forza Football AB needs to rely on its users for providing correct lineups for teams on their match-

day, the organization will put itself at risk since money, time and the organization's reputation might be at stake depending on the outcome of the situation. If the organization comes to the conclusion that a user cannot be trusted, no interaction will occur between these entities, meaning that none of the parties will gain anything, leading to a payoff of zero for both. However, if the organization would find the user to be trustworthy and the user actually would meet the desired outcome, both parties would gain something from the trust relation, leading to a payoff of one, respectively. The organization would gain what they expected, an accurate lineup for a team within the specified time frame. What the user would gain in such an outcome, is in this particular case, getting its reported lineup displayed in Forza Football together with the name that the user entered in conjunction with sending in the lineup. But if the user, on the other hand, fails to meet the organization's expectations, for example by providing an incorrect lineup or an invalid source, even though the organization trusted the user, the organization would loose time, money etc. that has been put at risk for this trust relation. For the user on the other hand, there is not that much to gain by creating a distrust relation to the organization. What potentially could be gained for the user is experience and learnings.

One of the most important aspects of trust is that it involves risks. Taking risks means diving into the unknown. In *Trust: Making and Breaking Cooperative Relations*, Niklas Luhmann makes a distinction between trust and familiarity [18]. According to Luhmann, trust is a matter of risk taking, while familiarity is an unavoidable factor since familiar things cross our lives on a daily basis. Familiarity lacks complexity, since what is familiar to us is already known, meaning that familiarity is a relatively simple aspect [18]. Since trust is a function of decreasing the complexity and risks, it means that there is no room for trust within familiarity [24]. However, Luhmann also points out that familiarity cannot be neglected within the concept of trust, since trust is an occurrence within a familiar world where changes are at hand, causing an impact on the possibility of creating trust relations [18]. These statements are rather contradictory, saying that trust cannot take place within a familiar world, while familiarity cannot be neglected when discussing the concept of trust. The meaning of these contradictory statements is that trust regards the future, but the future cannot exist without the past, meaning that familiarity is a precondition for trust. Trust is not something that can be established at a first glance. There is a need for past experiences enabling us to make the decision regarding whether or not to trust another entity and thus dive into the unknown - the unfamiliar world - not knowing the outcome with certainty [24].

Besides the prerequisite of familiarity for establishing trust, Luhmann further points out that self-assurance is another important aspect. Having self-assurance means having an inner security, which enables the entity to anticipate potential negative outcomes with tranquility, without immediately acting on a conceivable possibility.

"*Trust can come about if these internal reduction mechanisms are stabilized in such a way that they complement the environmental reduction and thus are in a position to reinforce it at critical points. In other words, bestowing trust is made possible and*

easy by the fact that the trusting system has inner resources available which are not structurally tied up, and which, in the case of a disappointment of trust, can be put into action and take over the burden of the reduction of complexity and the solving of problems."

This is a statement taken from *Trust and Power*, a book written by Luhmann in 1979 [24]. And according to this, Luhmann means that the presence of self-confidence decreases the complexity existing with trust, giving the trustor some certainty of the outcome. Besides bringing up the prerequisites for trust and the involvement of risks, Luhmann further states that trust can only exist where the possible outcome has a greater damage than advantage. This due to that, otherwise, the potential risks involved would be within acceptable limits, making the decision and actions easy for the entity. In those situations where an action or decision would make someone feel regret due to a negative outcome, is where trust is required [18].

Further work on trust has been done by D. H. McKnight and N. L. Chervany in their technical report *The Meaning of Trust*, where they, among other things, have taken the most common and important aspects of trust and defined six trust constructs: Trusting Intention, Trusting Behaviour, Trusting Beliefs, System Trust, Dispositional Trust and Situational Decision to Trust [9]. The Trusting Intention is "*the extent to which one party is willing to depend on the other party in a given situation with a feeling of relatively security, even though negative consequences are possible.*" [9]. The trusting intention is both situation-specific and intentional, since one entity is willing to depend on another entity in a given situation. One entity A having a Trusting Intention towards another entity B can reach a so called **reciprocal dyadic trusting intention relationship**, which means that B also has a trusting intention towards entity A. There are certainly even higher level combinations with more entities involved. Due to the large complexity of making an analysis with numerous entities involved, McKnight and Chervany define the Trust Intention on an individual level, meaning that it is one-way directional, having one entity relying on another entity, but not vice versa. There are five elements that constitute the trusting intention: **Potential negative consequences, Dependence, Feelings of security, Situation-specific context and Lack of reliance on control**, which all previously have been discussed.

The Trusting Behaviour is "*the extent to which one person voluntarily depends on another person in a specific situation with a feeling of relative security, even though negative consequences are possible.*" [9]. What distinguishes the trusting behaviour from the Trusting Intention is the behavioral term *depends* as opposed to the intentional (cognitive-based) construct *willingness* to depend. When depending on another person, one gives the person a certain power over oneself, meaning being placed in a situation of risk. Trusting Behaviour thus implies that risks have been accepted. The entity depending on another entity is thus placed in a situation with low or no control at all.

Trusting beliefs regards the extent to which one entity believes that another entity

is trustworthy. Viewing a trustee as trustworthy means that the trustee is able and willing to act in the trustor's best interest. The trusting beliefs can be narrowed down into different kinds of constructs, and McKnight and Chervany have chosen the following four:

1. **Benevolence** - having another entity's best interest in mind, and thus acting accordingly.
2. **Honesty** - speaking the truth, completing and fulfilling what has been agreed upon.
3. **Competence** - having the ability and right skills to complete what is expected by the other entity.
4. **Predictability** - consistent actions leading to that another entity can forecast upcoming actions in a given situation.

The combination of these four beliefs provides a solid foundation both for trusting intention and trusting behaviour. As McKnight and Chervany put it "*That is, if one is consistently (predictably) proven to be willing (benevolent) and able (competent) to serve the trustor's interests in an (honest) manner, then one is worthy of trust indeed.*". With these beliefs, an entity has established cognitive constructs called Trusting Beliefs finding their expressions in the Trusting Intentions being acted upon within the Trusting Behaviour.

System trust is "*the extent to which one believes that proper impersonal structures are in place to enable one to anticipate a successful future endeavor*". McKnight and Chervany differentiate the impersonal structures into: structural assurance and situational normality. The former one regards things such as regulations or contracts. The latter one refers to the entity's or other entities role within a particular situation. Trust Intention is supported by the System Trust due to the fact that the latter provides structural assurance safeguards, making it safe for the entity to depend on another entity and thus taking risks. In social systems for example, laws and their enforcement are adapted in order to make these situations controllable, providing these assurance safeguards.

Dispositional Trust regards the extent to which an entity has a consistent tendency to "*trust across a broad spectrum of situations and persons.*". There are two different kinds of directions of Dispositional Trust. One of them is known as Belief-in-People which means that an entity in general finds other entities to be trustworthy and thus trusts other entities in almost any situation. The other one, called Trusting Stance, regards those entities that believe that one may obtain a better outcome by trusting other entities, and thus generally trusts other entities. The Dispositional Trust construct becomes directed to people, meaning that people should be trusted in general, still within a context and familiarity, when Trusting Stance and Belief-in-People are combined. Belief-in-People, believing that other people are trustworthy in general, means that one has Trusting Beliefs, which eventually turns into Trusting Intention. As for Trusting Stance, believing that trusting other entities leads to a better outcome, encourages the entity to be dependent on others, meaning that

Trusting Stance directly leads to Trusting Intention, and does thus not affect the Trusting Beliefs.

The final one of the six trust constructs presented by McKnight and Chervany is Situational Decision of Trust, which regards the extent to which an entity aims at depending on a non-specific entity in a particular situation. This means that an entity has decided beforehand to always trust within a particular situation, no matter the entities involved, since trusting in that particular situation provides benefits that outweigh the potential negative outcome. Situational Decision to Trust does not support Trusting Beliefs since it is not concerned with the trustworthiness of another entity. However, since it is encouraged to be willing to depend on another entity in a particular situation, it does directly support Trusting Intention.

At this point, a general viewpoint of the concept of trust has been provided. What constitutes trust and what initiates it, and what problems exist both beforehand but also after a decision has been made to trust another entity, has also been provided. Even though the concept of trust is a wide one, consisting of numerous definitions by various researchers, there are common characteristics representing the concept of trust, which have been presented in Chapter 2. The concept of trust is most certainly a complex one, consisting of more aspects than one might imagine at a single glance. Trusting another entity means depending on that the entity will make decisions and take actions that are in ones best interest. This dependence between two entities is a matter of risk-taking since the trusting entity cannot affect the outcome with external enforcement. Assessing the trustworthiness of the actions completed by the entity is a hard task at hand, and in the following section, trust management will be presented.

3.2 Managing Trust Using Trust Management Systems

The traditional way of establishing trust in physical form by meetings and interactions between two entities, cannot be applied in the same manner when it comes to computer systems and applications. When information for a system is to be retrieved through crowdsourcing via external entities, the assessment of a system's trustworthiness becomes relatively difficult, since the information is provided by the crowd. And when the crowd is used for such a purpose, from the system creators perspective, there is a need for a trust relation between the creators and the crowd, where the creators need to trust in the crowd to provide information that is of gain to the creators, as discussed in Section 3.1. This also gives the organization behind the system numerous challenges to tackle in order to create customer relationships which induce trust. Managing trust within an environment where the online interaction between the organization and the crowds is required, has a focus on understanding and facilitating the trust between these involved entities. One of the purposes of trust management is to stimulate the external entities' and the organization's positive opinions towards the online environment's possibilities for

interactions. For an organization to assess the trustworthiness of the information retrieved through crowdsourcing using external entities, the organization's system is in a great need of a strategy and methodology enabling the organization to decide the trustworthiness of remote entities. The organization is also in need of creating customer relationships which induce trust, and thus, the trustworthy entities are also in need of methodologies enabling them to be acknowledged for the work they provide. Being acknowledged means receiving some sort of recognition for taking part in an organization's crowdsourcing project. Systems applying such methodologies are known as trust management systems. A system of this kind computes a certain trust value in order to enable the possibility of decision making in an automated manner. A trust value is typically a binary value between 0-1, indicating how trustworthy the user, and thus also the provided information by that particular user, may be. Since trust evolves over time, so does the trust value. In Jøsang's article, Can we manage trust? [11], trust management for online environments is defined as follows:

"*The activity of creating systems and methods that allow relying parties to make assessments and decisions regarding the dependability of potential transactions involving risk, and that also allow players and system owners to increase and correctly represent the reliability of themselves and their systems.*"

The definitions of trust and trust management have lead to computational models that focus on different aspects of trust management, evaluating the trustworthiness of entities and the information provided by them. In order for an entity to evaluate the trustworthiness of another certain entity, these models use probabilistic, socio-cognitive or organizational techniques. When the trustworthiness of an entity has been determined, the entity is enabled to make trust-aware decisions regarding whether or not to interact with the given entity at a certain point in time.

There are four specializations of trust management systems; direct trust evaluation, reputation-based trust evaluation, socio-cognitive trust evaluation and organization trust evaluation, which will be presented in their respective section [25]. The evaluation method relevant for this study is Direct trust evaluation, where the reasons behind the choice are given in Section 3.2.6. The other evaluation methods are still presented for the sake of context, and it is up to the reader whether or not to read them.

3.2.1 Direct trust evaluation

There are numerous ways of establishing trust between entities, and one possible way is by the observation of previous interactions with entities. These interactions consist both of positive and negative experiences between the two entities, as well as the timespan of each interaction. Modelling trust between entities can be done by viewing the interaction risk as the probability of potentially being misled by the interaction entity. This kind of probability can evolve over time by a trustor who looks at previous interactions with the trustee. These previous interactions become

direct evidence for the trustor to decide the trustworthiness of the trustee. The most common evaluation of a trustee entity's trustworthiness is defined within a certain context. The context usually consists of previous interactions that have occurred within that particular context. This makes the trustworthiness of an entity to only be valid within that particular context, which makes it necessary for the trustor to distinguish the relevant information for making proper interaction decision at a particular moment [25].

One instance of the direct trust evaluation model, was presented by Wang and Singh in 2007, where the presence of the uncertainty in trust evaluation was quantified [23]. An example of a potential scenario is when trustor A only has had interactions at two occasions with a trustee B, where the outcome in both cases has been successful, whereas another trustor, C has had 50 interactions with B, where only half of them have been successful. Which of these two gives enough evidence to assess the trustworthiness of trustee B, with least uncertainty? In their article, Wang and Singh tackled this problem by providing a function that calculates the uncertainty in a set of trust evidence, which are based on the distribution of positive and negative outcomes. The method provides a certainty value within a range [0, 1], which is based in statistical inference. The value 0 represents the highest value of uncertainty on the scale, while 1 indicates the opposite. As previously mentioned, in Section 3.1, the certainty of a trustee will be high if: 1. The amount of trust evidence is high; and 2. The conflicts between positive and negative outcomes is low.

3.2.2 Reputation-based trust evaluation

A reputation-based trust model combines the direct and indirect interactions and events with an entity in order to estimate that entity's trustworthiness. Indirect interactions are between the trustee and a third party, who in turn determines a direct trust value and makes it available for the trustor to use. One example of such a case is for example the reputation system, TrustPilot, where entities rate their interactions and events with an organization [26]. The overall reputation of the entity being rated will be the sum of all the ratings over a certain timespan. For reputation based trust evaluation systems there are general design guidelines provided [27].

1. "*The system should be self-policing. That is, the shared ethics of the user population are defined and enforced by the peers themselves and not by some central authority.*
2. *The system should maintain anonymity. That is, a peer's reputation should be associated with an opaque identifier (such as the peer's Gnutella username) rather than with an externally associated identity (such as a peer's IP address).*
3. *The system should not assign any profit to newcomers. That is, reputation should be obtained by consistent good behavior through several transactions, and it should not be advantageous for malicious peers with poor reputations to continuously change their opaque identifiers to obtain newcomers status.*
4. *The system should have minimal overhead in terms of computation, infrastructure, storage, and message complexity.*

5. *The system should be robust to malicious collectives of peers who know one another and attempt to collectively subvert the system.*"

Another example of a reputation-based trust evaluation model is the Beta Reputation System [28]. This system uses the beta probability in order to gather feedback from previous interactions and events that have occurred between two entities in order to derive a repetitive value.

Yet another example of a system relying on reputation-based trust evaluation, is the EigenTrust algorithm. Within this model, each entity receives a global trust value based on previous actions. This global reputation is given by the local trust value, weighted by the global reputation. Both values are given by other entities [27].

3.2.3 Socio-cognitive trust evaluation

The socio-cognitive trust evaluation model is concerned with information regarding the entities' internal properties together with external properties which might affect the entities' behavior in future interactions [29]. Internal properties of an entity can be concerned with ability/competence (which includes knowledge and self-confidence with cognitive agents), and disposition (which includes willingness, persistence, engagement, etc. with cognitive agents). The evaluation of the entities external properties is typically not concerned with the entity itself, but rather about belief in the entities' ability to recognize, exploit and create opportunities [30].

One model proposed for the socio-cognitive trust evaluation was given by Falcone and Castelfranchi in their article Socio-Cognitive Model of Trust [30]. The model is concerned with an entities mental state of trust when it comes to cognitive terms such as beliefs and goals. According to the authors, trust can take three forms at the same time:

1. "*A mere mental attitude (prediction and evaluation) toward another agent, a simple disposition;*
2. *A decision to rely upon the other, i.e., an intention to delegate and to trust, which makes the trustier "vulnerable".*
3. *A behavior, i.e., the intentional act of trusting, and the consequent relation between the trustier and the trustee.*"

All of these become different sets of cognitive parts within the trustor's mind. In the model proposed by the authors, an agent can only trust another agent if the former agent has both goals and beliefs, and thus the trustor has to be a cognitive agent. However, the trustee does not have to be a cognitive agent, but could rather be a system. For an agent X that trusts another agent Y with regard to Y's behavior/actions a for the relevant goal g gives:

- "*X is the (relying) agent, who feels trust; it is a cognitive agent endowed with internal explicit goals and beliefs (the trustier)*

- *Y is the agent or entity that is trusted (the trustee)*
- *X trusts Y about g/a and for g/a.*"

With these concepts (trust disposition, decision to trust, and trusting behavior), the authors state that an agent trusts another agent with regard to a given context. This context is concerned with the agent's motives, will, needs, objectives, just to mention a couple of examples. The relation between the three given concepts is that: agent X finds the actions of agent Y to be useful (trust disposition), X has made a decision of trusting agent Y's action (decision of trust), which means that agent X could come to hand out (act of trusting) action/goals to agent Y [30].

3.2.4 Organizational trust evaluation

The organizational trust evaluation model is typically concerned with establishing trust between contracting agents. These agents represent either an organization or a particular individual, negotiating contracts of goods and services and their deliveries. In order to enable automation of these transactions, it is essential for the organization to have some kind of insurance for preventing incorrect behavior. Systems managing this kind of trust combine their organizational structure together with their system. In order to establish such an approach there needs to be a trusted third-party entity, mediating the interactions and supervising the transactions between the contracting agents [25].

A representative example of an organizational trust evaluation approach is one of the frameworks available proposed by Kollingbaum and Norman in their article [31]. The framework consists of three different elements: 1) the transactional organization structure consisting of three roles (the addressee, the counter-party and the authority), 2) a contract specification language, and 3) contract management protocols developed by using the results from 2). The framework furthermore consists of three different roles, which is the addressee, counter-party and authority. In order to enable transactions between the entities involved, an agent is required to register with the authority, agree on terms for the contract with the agents and complete the required work stated in the contract, while the authority acts as a supervisor.

3.2.5 Trust-aware decision making

When making trust-aware decisions regarding whether or not an entity is trustworthy enough to interact with, the approaches at hand can be divided into two broad classifications: static (greedy) and dynamic. These classifications, in conjunction with one of the trust evaluation models previously presented, are used for choosing the most trustworthy entities for the different interactions, as Figure 3.2 presents, [4], [25]. The reason for the third-party testimony being grayed out in the figure, in comparison to the rest, is due to that it is not relevant for this study. The static approach normally uses more simple rules, while the latter classifications assess changing conditions within the system's environment in order to find a balance between the utilization of entities known to be trustworthy with the utilization of

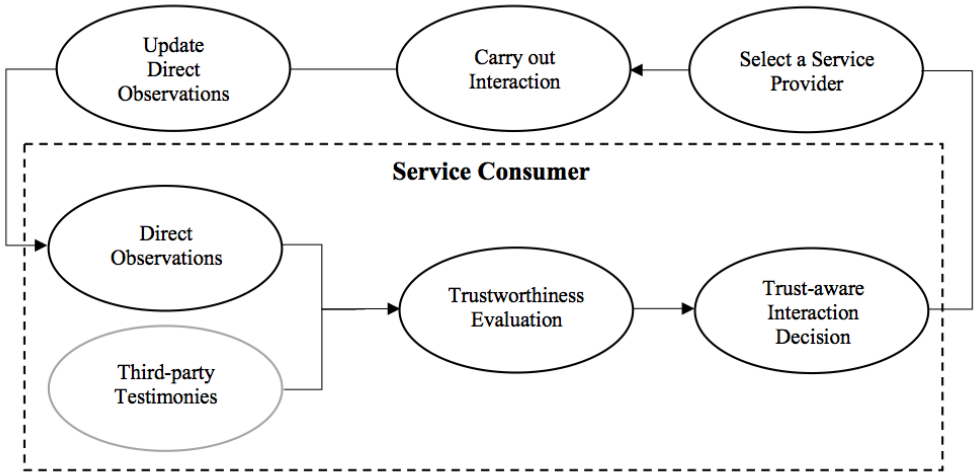


Figure 3.2: Model of typical trust-aware interaction decision making process. Model adapted from [4]

what potentially may be better alternatives. The trust in another entity can change depending on **direct trust**, which is based on an entity's personal experience and interaction with the entity, **indirect trust**, which also is referred to as reputation, is based on an entity's experience with a third party, and **self-trust**, based on an entity's advertising of oneself. The direct trust corresponds to the direct observations in Figure 3.2, while third-party testimonies is adapted by indirect trust.

Within the greedy approach, the trusting entity explores a trustee entity with a reputation through a supporting systems, such as the entity's social network or the other entities recommendations about that entity, or by using randomized exploration. One of the trust evaluation models previously presented is used in order to evaluate the reputation value retrieved for the entity, in order to find out which entity has the largest reputation. This determines which entity will be chosen for an interaction. When concerning only one trusting entity, this entity needs to make sure to choose the best possible option in order maximize its own long term well-being and success [4], [25]. This kind of approach is adopted mostly within computational trust literature, such as in the Beta-Reputation system [28], the article provided by Yu and Singh, *Searching Social Networks* [32] and *Trust-based Agent Community for Collaborative Recommendation* by Jianshu Weng, Chunyan Miao, Angela Goh Zhiqi Shen and Robert Gay [33].

As for the dynamic approach, there are a couple of approaches at hand, but not as many as for the greedy approach. On of them is presented by Victor Muñoz, Javier Murillo, Beatriz López, and Didac Busquets in *Strategies for Exploiting Trust Models in Competitive Multiagent Systems* [34]. In this article, the authors state that direct, indirect and self-trust all need to be factors part of the trust model, in order for the trusting entity to perform well. The decision-making process, consists of the resolution of exploration against the exploitation problem. The reason for

this is due to that the entity needs to decide if an interaction is to take place with an entity already known or with an unknown entity, opening up for the possibility of discovering even better information providers. If the entity only would come to focus on exploitation, it may provide good short-term results, but bad long term ones, whilst exploration gives the entity the possibility of discovering better entities to interact with, or substitute entities for when one of the usual entities fails, which is a rather common case when working in dynamic environments [34]. Also, exploration provides the trusting entity with adaption to faster changes, since if an existing interaction entity diminishes, the trusting entity can easily and fast change to another interaction entity. The authors base their trust model on the trusting entity's knowledge degrees about another entity, and use this in order to determine which entity is most suitable for the interaction in question. This knowledge degree is constituted by historical direct interaction experiences with the entity to be trusted, third-party testimonies and the entity's own reported trustworthiness. The authors state that a trust model can be anything that consists of the following:

1. "*The direct, indirect and self trust of the agents, for each of the services that they offer, with a normalizable value between 0 and 1.*
2. *The knowledge degree of a provider agent about a service (based only on direct trust). This is a value (normalizable between 0 and 1) that represents how much the agent has directly interacted with a provider, with 0 meaning that the provider is totally unknown (the agent has never directly interacted with that provider), and 1 totally known.*" [34].

In the case of Forza Reporter, the only trust evidence that the organization Forza Football AB has to go on, in order to determine which entities are trustworthy enough for providing correct lineups, is the direct interaction that the organization has with its users. Meaning, that trust model most suitable for Forza Reporter is the one based only on direct trust.

Since the decision-making process consists of the resolution of exploration against exploitation, it is of great importance for the trustor to decide how much that should be spent on exploring new unknown entities, and how much on exploiting already known ones. For enabling this, the authors have grouped the different entities into four categories, based on the information existing in the trust model regarding a certain task. The four groups are:

1. "*Group TK (Totally Known agents): The agents with a knowledge degree equal to 1. The trust model says that they are well-known providers since we have interacted with them many times.*
2. *Group PK (Partially Known agents): Agents with a knowledge degree lower than 1 and greater than 0. These providers would probably become future providers in the event any of the existing ones failed.*
3. *Group AU (Almost Unknown agents): Agents with a knowledge degree equal to 0. These are the agents without any direct interaction, but for which there is indirect or self information.*

4. *Group TU (Totally Unknown agents): These are the agents without any information about, either direct, indirect or self.*" [34].

This metric is normalized within a range of [0,1], where 1 represents "completely known" and 0 represents "no direct interaction experience". In the local data available for the trusting entity, the entities to be trusted are organized into four groups, based on their knowledge degree metric. If the number of entities with a reputation value higher than a threshold predefined by the trusting entity, in the most well known group, the trusting entity will only select among entities within this group for an interaction. Otherwise, exploration of the entities within the other groups will take place in order to increase their knowledge degree and promote these entities into higher order groups. [34]

3.2.6 Evaluation of the trust method for this study

The four different specializations of trust management systems that have been provided are all suitable in different contexts. In order to make a decision regarding which of these is the most suitable for the purpose of this study, the method of exclusion was used. Since this study does not concern any contracting agents and has no need for a third-party entity mediating the interactions between contracting agents, the organizational trust evaluation method can be excluded.

The Socio-cognitive trust evaluation model is concerned with two types of information regarding an entity: the internal and external properties. The latter case employs any kind of external factors that may affect the entity's future behaviour, meaning the belief in the entity's ability to recognize, exploit and create opportunities. Since the ability to provide correct lineups is a factor external to the entity, which affects the trustor's expectation of the trustee's future behaviour, and since this study's trustworthiness evaluation of an entity and the automated decision making thereafter will not be concerned with any kind of external factors that may affect the entity's future behaviour, the Socio-cognitive trust evaluation model may also be excluded.

This therefore leaves two evaluation models as potential candidates for this study: the Direct Trust Evaluation and the Reputation-based Trust Evaluation. The Reputation-based trust model combines the direct and indirect interactions and events with an entity in order to estimate that entity's trustworthiness. The latter case regards the entity's direct interactions with a third entity, who depicts this interaction between these entities. The Direct Trust Evaluation model on the other hand, completely relies on previous interactions with an entity, which consists both of positive and negative experiences between these two entities, as well as the timespan of each interaction. This kind of trust evaluation can evolve over time by a trustor entity who looks at previous interactions with the trustee entity, within a given context.

For this study, there are two entities: the organization Forza Football AB who de-

cides the trustworthiness of the second entity, the users. Since there is no third entity who depicts the interaction between these two entities in order to estimate the user's trustworthiness, there is therefore also a lack of reputation. That means that the Reputation-based trust model is not an appropriate choice for this study. Therefore, since the trust value will be given to a user based on all previous performances within a certain timespan in a given context, the type that will be used within this project is the Direct Trust evaluation model.

4

Crowdsourced Information Retrieval

The understanding of the concept of trust and the emergence of trust management and systems managing trust has opened up new possibilities for organizations and their systems. Taking advantage of this has enabled entities to widen the availability of information by retrieving it through crowdsourcing. Crowdsourcing is a concept first introduced by Jeff Howe in his article *The Rise of Crowdsourcing*, in which he defines crowdsourcing as an act of sharing information, where an organization outsources a function of pre-defined tasks to entities external to the organization [35]. This means that entities solving the problem have a limited freedom to operate. Howe defines Crowdsourcing, more precisely, as: "*Simply defined, crowdsourcing represents the act of a company or institution taking a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call. This can take the form of peer-production (when the job is performed collaboratively), but is also often undertaken by sole individuals. The crucial prerequisite is the use of the open call format and the large network of potential labourers*". As for Forza Football AB, the organization uses Forza Reporter in order to crowdsource data from their users, which in turn can be displayed in their live-score application Forza Football. The task is currently restricted only to lineups, where the crowd is expected to provide lineups for teams.

An addition to Howe's definition of crowdsourcing was made by Yochai Benkler. According to Benkler, the degree of uncertainty within a crowdsourcing project is very low since the tasks are predefined. Furthermore, the importance of the human knowledge input is also extremely low, as well as the degree of how formalizable, explicit, and routine the human knowledge actually is. In addition to this, crowdsourcing is seen as a great tool when cost reductions are the main matter, but not as suitable when a resource is to be explored and when the opportunity space needs to be wide [36].

Continued, Joseph Feller states that "*The acquisition of innovation capability (i.e. knowledge and skills) through direct or mediated means is a form of crowdsourcing.*". It is also pointed out that crowdsourcing has a great potential when open innovation strategies need to be implemented and is the centre when it comes to design of innovation intermediaries [37].

Since the term crowdsourcing emerged, numerous studies have continued on Jeff

Howe's work, trying to define it even further, providing possibilities and limitations of it. Another example is the book *Crowdsourcing*, written by Daren C. Brabham [38]. Brabham defines crowdsourcing within the online environment, where the online community referred to as the crowd, receives the opportunity of solving different kinds of crowdsourcing activities provided by an organization. Gathering a crowd like this for problem solving, under the right conditions, can come to provide large benefits and better outcomes for the organization than individual experts within the field would be able to do. Crowdsourcing is a matter of cooperation, teamwork and creativity.

Another definition of the term Crowdsourcing, based on a systematic analysis of numerous scholarly literature's on Crowdsourcing, was completed by Enrique Estellés-Arolas and Fernando González-Ladrónde-Guevara and documented in their article *Towards an integrating crowdsourcing definition*. The key items they found setting the basis for crowdsourcing are:

1. "*an organization that has a task it needs performed,*
2. *a community (crowd) that is willing to perform the task voluntarily,*
3. *an online environment that allows the work to take place and the community to interact with the organization, and*
4. *mutual benefit for the organization and the community.*" [39].

In order for mutual benefits to be possible for both the organization and the community, it is of great importance that an interplay between the two parties exists, enabling co-creative efforts from both. Having a control of the creation of ideas, making sure that it is a shared collaboration between the organization and the crowd, will maximizing the benefits for both parties. However, when this control is more on the organization's side, the crowd will become "one in the crowd", a pawn in the organization's goals, causing the larger benefits for the organization. If the control, on the other hand, is on the crowds side, the organization will become fortuitous to the crowd's work, merely working as a platform on which the crowd can build on using their own strategic goals. In this case, the benefits would be larger for the crowd. An example of such a case is Wikipedia, which merely is a platform where the crowd provides any kind of information. The information is added, deleted and edited by the crowd [38].

Adapting crowdsourcing within an organization means externalizing the process of problem-solving to the crowd, rather than solving it internally. This approach of problem-solving leads to a large and diverse set of skills, ideas and information flow used for solving the problem. What makes crowdsourcing possible is both technical and conceptual conditions. The technical aspect concerns technologies and the Internet that act as props for crowdsourcing applications. As for the conceptual aspect, "*crowdsourcing can be explained through the processes of problem solving and innovation as well as through the group phenomena of collective intelligence and the wisdom of crowds.*" [38]. In order for crowdsourcing to even be initiated, an organization is required to define a task that needs to be performed by the crowd,

who in a collaborative manner need to solve the problems at hand. Problem-solving is a key concept within crowdsourcing, since it allows the organization to open up the problem to the online community in order to solve the problems the organization currently is confronting. According to Kevin Dunbar, there are four components that constitute problem solving. First and foremost, there has to be an initial state, which refers to a persons initial state of knowledge towards the problem. The second component is the goal state, referring to what the person wants to achieve. The next component contains the actions taken by the person in order to reach the goal state. Finally, the fourth component is the actual environment where the actions take place. The environment is constituted by any physical aspects that may directly or indirectly affect the direction of the problem solving [40].

4.1 Commonly mistaken for crowdsourcing

There are numerous systems, applications and platforms that commonly are considered to be crowdsourcing, while they technically are not. One such example is Open-source projects. Open-source projects are a collective collaboration between individuals who produce a resource on their own terms, in their own format with the aim of reaching their own goals. These individuals voluntarily make contributions to the resource, by making improvements to it, adding new features, fixing bugs etc. Once a change to the resource has been complete, it becomes available to the community, free of charge. The biggest reason for why open-source projects are not considered to be crowdsourcing is due to its lack of top-down management, meaning that the project is lead by the community and not by any particular organization. Open-source projects are considered to be self-organizing collaborations between peers within the community, working towards their own common goal, and thus lack any kind of management steering the project. However, in recent years it has become rather common for open-projects to adapt top-down management processes, having someone steering the design and features of the projects. Despite this, open-source projects commonly steer away from the traditional way of developing resources, where an organization sets the goals and tasks, and is thus not considered to be crowdsourcing [38].

Another concept normally confused with Crowdsourcing is the term Common-Based Peer Production, introduced by Yochai Benkler. In Benkler's article *Peer Production and Cooperation*, he defines peer production as an organizational innovation where organizations completely rely on external entities, which in this case usually are users, as content providers and creators. He also further describes peer production as being an organizational innovation over three dimensions: "*(a) decentralization of conception and execution of problems and solutions, (b) harnessing diverse motivations, and (c) separation of governance and management from property and contract*" [36].

Peer production can easily get mixed up with the term crowdsourcing [5]. According to Benkler, there is a distinct difference between these two. Benkler puts these innovation types into perspective by having a three-dimensional graph; "*(a) the degree of*

uncertainty in the project space, (b) the degree to which the human knowledge input important, as well as the degree to which it is formalizable, explicit, and routine as opposed to tacit, intuitive, or creative, and (c) the degree of capital concentration required to execute the project" [36]. Crowdsourcing belongs to the lowest degree, in all three dimensions. This means that tasks are predefined in crowdsourcing, which gives a certainty for the organization, but at the same time limits the participants as for what they can do with the project. The limits existing in crowdsourcing do not "*restructure innovation, learning, and adaptation for the organization or the task, and does not generally harness any new motivational vectors beyond standard hedonic gains*" as Benkler puts it. Whereas in peer production, the participants have a larger freedom to operate and the tasks at hand are not controlled in the same manner as in crowdsourcing, having a top-down directive for what can be done by the participants. Peer productions thus belongs to the highest degree in these three given dimensions.

4.2 The use of crowdsourcing

In order for an organization to successfully solve their problems using crowdsourcing, the problem at hand needs to be properly and detailed enough defined, and the correct crowd needs to be chosen. Researchers have by different means categorized crowdsourcing cases within different kinds of typologies, some based on the problems being addressed and some based on the type of crowd required. One example is the typology based on six different crowds, proposed by Nicholas Carr. He means that the crowds available for solving problems using crowdsourcing are: social-production crowds, averaging crowds, data-mine crowds, networking crowds, transactional crowds, and event crowds [38]. This kind of typology lays focus on the crowd´s abilities and their strength to work together, but does not provide any guidance for the organization regarding how to use crowdsourcing in an advantageous way.

Eric Martineau presented a four-type-typology in his master thesis, *A Typology of Crowdsourcing Participation Styles*. This typology is based on two criteria: "*level of engagement and actions posed as part of crowdsourcing*" [41]. The typologies are: **Communals** - who constitute characteristics such as "*consumer empowerment, the need for recognition and incorporation of the brand into one's self*". The participants within this group develop social capital by participating on site. **Utilizers** - who get motivated to complete the different tasks by getting recognition through materialistic or symbolic benefits. The reason for completing the tasks within crowdsourcing is for self-expression or to improve oneself. **Aspirers** - who have a rather limited participation, since they are not content providers for the problem solving, but rather share their opinions and provide ratings about already created content, in order to help shaping the final version of the product. As opposed to the two other groups, this group does not only use their participation for self-expression, but also the consumption of the product. **Lurkers** - also known as observers, barely have any participation within the actual crowdsourcing project, and they do not invest many resources in it. They typically buy the end product, but are not concerned

with the process behind the creation of the product.

Another suggestion, by dividing crowdsourcing into four different strategies was made by Jeff Howe in his book *Crowdsourcing: Why the Power of the Crowd Is Driving the Future of Business*. Howe's way of dividing crowdsourcing into four strategies, focuses on the way applications are functioning. The four strategies are: crowd wisdom - which takes advantage of the crowds collective intelligence, crowd creation - taking advantage of the crowd for the development of the product that is to be sold, crowd voting - letting the crowd go through things and vote, and crowd funding - taking advantage of the collective resources of the crowd [42].

In 2011, Andrea Wiggins and Kevin Crowston wrote an article *From Conservation to Crowdsourcing: A Typology of Citizen Science*, where they analyzed different kinds of citizen science projects in order to find typologies. The typologies found are based on the organizations purpose for using the citizens. There were five typologies identified: Action, Conservation, Investigation, Virtual, and Education [43]. Action-oriented citizen science projects are normally planned and completed by the citizens. Most of the projects within this category adapt participatory action research approaches and has quite a focus on the physical places for where the volunteer participation takes place. Conversation projects have a focus on educational goals or content and it is thus a matter of practicality and outreach as for the citizens engagement. Investigation projects focus on "*scientific research goals requiring data collection from the physical environment*". Even though education is not an explicit goal of projects within this category, it is typically a part of the project's purpose and educational materials are thus often provided. Furthermore, the tasks are structured in such a way that they support continuous learning. Virtual projects are ICT focused and do thus not concern any kind of physical elements, as oppose to the other projects. The final typology, Education, has an explicit goal focusing on education and outreach. The authors further divide this typology into two subdivisions: formal and informal learning opportunities. Furthermore, projects within this category normally support continuous learning, which is a shared feature with many projects within the Investigation category [43].

In his book, *Crowdsourcing*, Daren C. Brabham investigated previous work on typologies, finding numerous different ways of defining crowdsourcing. He came up with a four-type typology which is based on the problems being solved within the organization. The four typologies are: the knowledge-discovery and -management approach, the broadcast-search approach, the peer-vetted creative-production approach, and the distributed-human-intelligence tasking approach [38]. The knowledge-discovery and -management approach is best suitable to use when information is to be collected, organized and managed, which is the case of Forza Football. Furthermore, this approach is adaptive when collective resources are to be created. When an organization chooses this approach, they provide the crowd with the task of gathering information into a common format and place. The broadcast-search approach is adaptive for scientific problems, where ideas are to be developed, eventually leading to empirical solutions. Organizations choosing this approach will hand out tasks

to the crowd, who is responsible for solving empirical problems. The peer-vetted creative-production approach is suitable when the organization has design or aesthetics problems, and needs help from the crowd to come up with ideas of solutions. These solutions are normally influenced by the crowds taste or by the market support. The organization provides the crowd with the task of creating creative ideas and also selecting them. The final approach, distributed-human-intelligence tasking, puts weight on data analyses where the human intelligence is the priority, rather than computer analyses. These cases apply when the human intelligence is more effective or efficient than what a computer analysis would be. Within this approach, the crowd's task is to analyze large-scale information. Even though there are numerous definitions and typologies of crowdsourcing, it is of great importance that the organization defines the problem they need solved explicitly and detailed enough. Furthermore, in order for the task to be solved in the right manner, the organization needs to make sure that the correct crowd is chosen. For example, having a crowd consisting of experts at cars will not be that helpful for solving a problem for a football application. Therefore, the crowd needs to possess proper knowledge and skills for solving the problem, and should thus be carefully chosen accordingly.

4.3 Issues in crowdsourcing

Crowdsourcing may be the key to success, if used correctly. There are however numerous issues existing within crowdsourcing, which cautiously need to be taken into account and handled by the organization adapting it. Forza Football AB have been facing several issues since the start of adapting crowdsourcing as a part of the organization's mission. The issues brought up within this section are relevant for this study and Forza Football AB. Additional issues do exist, such as legal issues, problems with free speech and dissent, intellectual property and copyright, and Ethical issues, just to mention a few. But since these issues are not relevant for Forza Football AB, they are therefore not relevant for this study and will thus not be further mentioned.

Within any crowdsourcing project, the participation of the crowd is motivated in some manner, may it be honor, experience, money or even physical objects. It is necessary for the organization to know what motivates the crowd's participation in order to make the crowdsourcing project as efficient and effective as possible. In their article, *Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being*, Richard M. Ryan and Edward L. Deci, write that motivation is a concern of energy, direction, persistence and equifinality [44]. When requiring others to make a certain action, such as in crowdsourcing, motivation becomes a highly valued aspect, since motivation is what makes people produce things. Understanding what motivates people is a hard task at hand, since different people are motivated by varied factors, which is influenced by different experiences and consequences. This is an issue that Forza Football AB have been experiencing, with users contacting the organization, where some simply want to receive recognition for providing lineups, whilst others expect even more, such as economical compensation. Different people evidently get motivated by different aspects. The motivation that a person possesses can either be influenced by strong external coer-

cion or by a particular activity valued by the person. The person may act directly from a personal commitment or due to a bribe or surveillance. The different ways of being motivated can be put into two categories: internal and external. When comparing between people who authentically (self-authored or endorsed) get motivated and those who are controlled by external forces, it becomes evident that those within the former one are more interested, confident and excited to complete the work they have been assigned, and therefore have an enhanced performance and creativity [44].

In their self-determination theory Richard M. Ryan and Edward L. Deci try to define different types of motivation, mainly differentiating the motivations of people between intrinsic and extrinsic. In their article, the authors create a spectrum where they put the different motivation types in order, in terms of the "*degree to which the motivations emanate from the self*", as the authors put it [44]. On the leftmost side is amotivation, which is the lack of an intent to act. Amotivation is a result from the fact that an activity is not valued enough, the competence for it is lacking or that it is not expected to give the desired outcome. To the rightmost side in the spectrum is the intrinsic motivation, which is defined as "*doing an activity for the inherent satisfaction of the activity itself.* Intrinsic motivation can only exist if the activity completed is within the persons intrinsic interest, having the appeal of novelty, challenge, or aesthetic value. In order to enhance the intrinsic motivation, researchers have found that social-contextual events, such as feedback and communication, that makes the person feel competent, are contributing factors, which is something that Forza Football AB are in a great need of. As mentioned in Chapter 2, Forza Football AB currently face a problem with a lack of feedback, which is one of the reasons for some users not being motivated enough to provide correct lineups or any lineups at all. The organization is thus in a great need of mechanisms that motivate their users, both in order to maintain the community of already existing users, but also to motivate new ones. Further facilitating factors, for making a person feeling motivated, are optimal challenges and positive performance feedback.

Between the amotivation and intrinsic motivation, lies the extrinsic motivation, which is defined as "*the performance of an activity in order to attain some separable outcome*" [44]. Extrinsic motivation can be further divided into four different types: External Regulation, Introjected Regulation, Identified Regulation and Integrated Regulation. Actions taking place within the External Regulation are performed in order to satisfy an external force or demand. Externally regulated behaviour is normally conceived as controlled. The Introjected Regulation is a somewhat controlled form, in which performances are done in order to retrieve self-control, attain ego enhancement or for internal rewards. The third form of extrinsic motivation is viewed to be more self-determined than the two former ones, in the sense that the extrinsic motivation comes from identification. The identification is something that is valued, such as a behavioral goal, leading to that the action becomes accepted as being personally important. The final type is integrated regulation, which occurs when all aspects have been evaluated and accepted by the individual, conforming to the individual's own values and needs. Even though this type of extrinsic motivation shares some qualities with the intrinsic motivation, it is not considered

to be a part of the intrinsic motivation due to that the actions are completed for reaching separable outcomes, as oppose to the aim of attaining inherent enjoyment in intrinsic motivations. When intrinsic and extrinsic motivators intersect, it is however normal that the participant's motivations shifts more to the extrinsic side (such as extrinsic rewards), leaving the intrinsic motivators undermined, leading to that the participant engages in activities both for intrinsic and extrinsic motivations.

Even though different individuals may be motivated by different means, researchers have found certain motivations to be common within crowdsourcing projects. The motivations are [38]:

- Money.
- Developing one's own skills.
- Establishing contacts in the form of other professionals.
- Tackling tough problem-solving.
- Socialize and establish friendships.
- Occupying oneself when being bored.
- Contributing to large scale projects.
- Sharing resources with others.
- Having fun and enjoying oneself.

It is evidently impossible, or at least a very hard task at hand, trying to satisfy each individual's motivations and needs. Even though it is necessary for the organization to know what motivates the crowd's participation in order to make the crowdsourcing project as efficient and effective as possible, the organization cannot stretch itself to the extent that they provide or promise resources they do not have. Many of the users in Forza Football's crowdsourcing project, Forza Reporter, have done it for self recognition, establishing contacts and developing their own skills. Getting acknowledged and getting ones name displayed within the application, for the reasons of being the fastest lineup provider, has been a win in itself for many of the users. However, as in any other crowdsourcing project, there have been those users who expect to get something more valuable in return, such as money. Since this, however, is not the organizations intent with Forza Reporter, and since they simply do not have the intention on providing such resources to its users in order to increase their motivation, the organization should not go beyond its own limits for satisfying these users. The organization thus needs mechanisms that satisfies the majority of the users, who are satisfied with acknowledgment and self-recognition, but at the same time provide informative feedback to the users letting them know what the purpose of the project is and what is in it for the users to participate. And by doing so, the organization will handle those users who expect more. Just as the users have their motivations for participating, and the organization is obligated to know what these motivations may be for the efficiency of the project, so does the organization also have. The organization has its own motivations, incentives and limits, and these cannot be overseen just for satisfying a couple of users among numerous others.

5

Trust Management Systems in Crowdsourcing

The great tools existing for retrieving information through crowdsourcing and the opportunities existing with gathering information in such a manner, has opened up for unprecedented information and data, useful for solving numerous problems at hand, such as the case of Forza Football. Crowdsourcing systems have been found to be powerful platforms for the utilization of the crowds' skills and knowledge. Nevertheless, as with any other open system, great opportunities such as these also open up for risks, such as the existence of malicious and selfish behaviour of users. When finding the correct crowd for crowdsourced information retrieval projects and mechanisms for motivating them for participating in the project, as discussed in Chapter 4, one of the greatest issues and obstacles arising in crowdsourcing projects is the assessment of the trustworthiness of the information provided, which relates to the range of correctness of the users taking part in the project [45]. In many of the crowdsourcing systems existing today, the participants in the crowdsourcing projects are normally anonymous and thus completely unknown to the organization behind the system, as in the case of Forza Football. In many of the crowdsourcing systems, there are also no resources provided for the organization to handle information of low quality. This case is solved by Forza Football, by requiring users to include an official source when reporting a lineup, and by manually reviewing each lineup, as mentioned in Section 2.1. In current systems adapting crowdsourcing, no particular way of analyzing the trustworthiness of the users and the information they provide, has been provided. In order to solve this issue, there has been a great need for trust management mechanisms in crowdsourcing. As presented in Section 3.2, the purpose of trust management is to identify the most trustworthy users to the best extent possible for a given task. Adapting such trust management mechanisms in crowdsourcing projects is a feasible way of distinguishing trustworthy users from those who are not, and also to tackle the issue of malicious users. As discussed in chapter 3, the trustworthiness of a trustee is the subjective probability by which that trustee is expected to perform a given action for which a trustor's welfare depends on. Thus, it is useful with trust-aware mechanisms that enhance the social welfare of an entire system adapting crowdsourcing [45].

One of the greatest challenges with integrating trust management models into crowdsourcing systems, is not only finding the most trustworthy users, but also knowing how to utilize these users. In the long run, users may terminate their participation in the crowdsourcing project. And if there is no way of retaining these users, the

organization behind the system will not reach their goals in time, leading to them turning away from the crowdsourcing system. Such a cycle affects the normal operations of the crowdsourcing system in a negative manner. In order to retain and optimize the social welfare of the crowdsourcing system, both these issues need to be solved. To work towards this, it is therefore recommended that trust management methods also consider the social welfare, and not only focus on the organization's benefit. In their conference paper, Han Yu and Chunyan Miao suggest a couple of research directions worth pursuing in order to reach these goals [4]. The first approach is to approximate centralized optimization. This is done by making use of the information already available in the system, where approximate centralized optimization might be of help for achieving fair treatment of the trustworthy users taking part in the crowdsourcing projects, and "*compute an approximately optimal solution in consideration of the complexity of the global optimization and time constraints*" [4]. The second approach is coordination, a mechanism for giving the organization an indication of when it is more profitable for the social welfare to engage less trustworthy users in the project. Learning is the third approach proposed by the authors. It is possible for the organization to estimate the level of workload, as for the waiting time and make adjustments to their strategies accordingly. The final approach proposed is negotiation. The organization and the users taking part in the crowdsourcing project may design interaction protocols between them, giving the two parties room for negotiating the terms of their relation.

These approaches are not meant to be comprehensive, but are rather meant to serve as guidance for those adapting trust management into their systems, providing an understanding of the issues existing when integrating trust management into crowdsourcing systems [4].

6

What Is Feedback And Why Is It Important

Feedback is of great importance, because it shows users that the information they provide actually is being used, or at least acknowledged. Furthermore, feedback is a key for giving users an incentive to do more and better, and the feeling of recognition is what many seek [1]. With the current state of Forza Reporter, with no feedback provided to users when reporting lineups, there is evidently a great need to integrate feedback into the current process, in order to provide incentives for users to report more and better lineups. Therefore within this chapter, definitions of feedback and how it can be used will be provided, along with reasons for why it can be such a powerful tool.

6.1 What feedback is and how to use it

Feedback is a process used for decreasing the gap between entities and is a very powerful tool if used correctly. When providing proper feedback, entities receive guidance in how to complete the task more accurately and faster. This is potentially an attribute of that continuous knowledge about the entities' performances is enabled by the feedback that is given. According to a study conducted and presented in "*Motivational Feedback in Crowdsourcing: a Case Study in Speech Transcription*", entities that receive feedback about their performances, tend to provide results with higher accuracy and also with a higher typing rate. [46]. The authors further state that gamification techniques, such as leader boards, levels and badges, can increase the enjoyment and make the process more playful. In their study, it appeared that the use of gamification increases entities' quantity and quality of work. However, gamification was also found to be effective and motivational only to a certain threshold, for which, if exceeded, the intrinsic motivation of entities, for completing a task correctly gets replaced by entities' endeavors for more achievement feedback. In order to keep entities' motivations on the right track, it is of greatest importance that a maximum safe level is found, meaning that the combination of several gamified elements might not be effective. When providing feedback on the quality of the data provided by entities, it may yield an increase in their overall quality, since it makes the entities put more effort into it. Furthermore, when entities receive recognition in such a manner that their contributions also are displayed to others, volunteerism increases.

In their article, *The Power of Feedback*, John Hattie and Helen Timperley provide a model of feedback used for enhancing learning. They define the purpose of feedback as a tool to "reduce discrepancies between current understandings/performances and a desired goal" [47]. These discrepancies can be reduced by having the two entities involved doing the following: **Entities receiving feedback:** Will either increase their effort and use more effective strategies, or abandon or lower their goals. Whilst **Entities providing feedback** will provide goals and challenges appropriate and specific enough, and assist the entity on the receiving end to reach the set goals by adapting proper strategies and feedback. When providing effective feedback, three questions need to be answered: *Where am I going? How am I going? Where to next?*. These feedback questions each work on the following four levels:

- **Task level**: How well a specific task has been performed by the entity.
- **Process level:** The process used and needed for best performance possible of the task.
- **Self-regulation level**: Monitoring and directing oneself.
- **Self level:** Personal evaluation and providing affects about the entity.

Providing feedback on task level, which also is referred to as outcome feedback, or knowledge of results, means providing information about the results, even though the outcome may be correct or incorrect. Feedback on task level may also include the depth and the quality of the work completed by the entity. This kind of information is typically either against explicit or implicit criteria. Furthermore, this kind of feedback may also regard the format of which the task has been completed in or even the need for further information. This means that outcome feedback does not provide any further details other than the achievement of the task [48], [47]. It has been found that feedback on task level comes in handy when misconceptions are to be corrected when the entity has provided too little information. In these cases, further instructions on how to complete a task is more powerful than feedback actually would be. The problem with feedback on task level is that it is task-specific, meaning that it most likely will not transfer to other tasks as well. It does most certainly provide a learning for the entity, for the task at hand, but does not provide as much learning as feedback on the tasks process level [47].

Feedback on the tasks process level provides the entity with information about the approach used to solve the task, the relationship between the results and the performance quality, and information about other strategies that potentially could have been used. There are entities that have the ability to take in outcome feedback and turn it into cognitive feedback, which is finding the binding component between a tasks characteristics and its processes together with the results. Cognitive feedback can be divided further into three different types: Task validity feedback, cognitive validity feedback and functional validity feedback. The task validity feedback provides information about the relation between a tasks achievements and cues. These kind of relations can regard for example principles within an experiment, or even based on historical data about the entity, gathered within a specific environment. The cognitive validity feedback regards the relationship between the cues and achieve-

ments. This kind of feedback provides the entity with information about how to perceive cues and assess how performance can be influenced by them. For the last type, the functional validity feedback, the entity receives information about relation between the entity's estimates of potential achievements and the actual outcome [48].

Self-regulation is a process in which the entity controls its own learning. The act of self-regulation may lead to that the entity seeks, accepts and acts upon feedback. Those entities that are effective learners, usually find a way of continuously figuring out if and when there is a need for more information and find appropriate strategies for receiving the feedback. Whilst less effective learners tend to find themselves being more dependent on external factors, meaning awaiting others to provide feedback rather than seeking it. Those entities who have self-efficacy, which is confidence in oneself as a learner, and confidence in that the information received in conjunction with feedback is worthwhile, tend to make an effort in both receiving and dealing with this kind of feedback. Thus, giving feedback about self-regulation is seen to be effective to the extent that it enhances self-efficacy [49].

In her book, *How to Give Effective Feedback to Your Students*, Susan M. Brookhart provides common characteristics for feedback [49]. Even though this paper is directed towards feedback to students, it can still be applied in other contexts as well, since the same feedback can be applied to any crowd. Furthermore, the points brought up by the author are relevant for the case of Forza Reporter and thus, this study. In order to make sure to give good and informative feedback, it is of great importance that the correct strategies and content is chosen. The strategies in feedback can vary in timing, amount, mode, audience, and the content in feedback can vary in focus, comparison, function, valence, clarity, specificity and tone.

6.1.1 Feedback strategies

The timing of feedback can vary in when it is given and how often that happens. In order for the timing of feedback to be good, the author recommends that feedback should be provided immediately, while it is still fresh in an entity's mind, letting the entity know if something is right or wrong. Furthermore, feedback should be provided as often as it is proven to be practical. The purpose of timing the feedback is to give feedback to the entity while the entity's mind is still set on the learning target and while there is still time for the entity to act on it.

Examples of good timing:

- Returning results of the data provided by the entity as soon as possible. Next day at the latest.
- Providing immediate feedback about an entity's misconception.

Examples of bad timing:

- Returning results of the data provided by the entity several days upon receiving.
- Error or misconception ignorance.

- Providing feedback first when there is no opportunity for improvement.

In order to decide what the right amount of feedback is, it requires consideration of the topic and its targets, the typical developmental learning progressions and each individual entity. These three need to be considered simultaneously in order to be able to asses the proper amount of feedback to give. For good amount of feedback, one can follow the Goldilock principle, which says "Not too much, not too little, but just right". The purpose of making sure to provide the proper amount of feedback is in order to give entities enough information to know what to do next, but not so much that the works is done for them. **Examples of good amount:** is when important learning targets are included within the feedback. It is also beneficial when providing feedback on at least as many strengths as potential weaknesses. **Examples of bad amount:** is when pinpointing every exact error and mistake, and when providing voluminous comments on poor-quality work, but not on good-quality.

The mode of feedback can take three various forms: oral, written or visual. For the purpose of this study, oral feedback is not possible since there is no such interaction with users. Therefore, the feedback in this study will only focus in written and visual feedback. Visual feedback should only be provided if there is an issue regarding "How to do something", or if an example needs to be provided.

The audience can be either individual or a group. Group feedback only works when an entire group of peers have reached the same outcome, and can thus receive the same feedback. However, in this study, each user is treated as their own individual, and it is therefore crucial to give feedback accordingly. An example of good feedback, is communicating with a specific entity, giving feedback about the entity's own performance.

6.1.2 Feedback content

The purpose of feedback focus is to provide the entity with the quality of the work achieved in relation to potential learning goals, make an effort in improving the entity's strategies and processes, increase self-efficacy by providing the relation between the work completed and the efforts made, and avoid personal comments.

Examples of good feedback focus:

- Providing information about both strength and weaknesses of a particular performance.
- Providing recommendations about certain strategies in order to foster achievement.
- Providing feedback, giving the entity a sense of that it is the entity who is completing the work.

Examples of bad feedback focus:

- Providing feedback that will bypass the entity.

- Providing criticizing feedback, without providing ways of improvement.
- Providing feedback with personal comments.

As for the comparison of the content, one entity's content can either be compared to criteria (criterion-referenced), other entities work (norm-referenced) or to the entity's own previous performances (self-referenced). The criterion-referenced feedback should only be given when feedback is to be provided about the entity's actual work. This kind of feedback helps the entity to decide its next-coming goal. The norm-referenced feedback should be provided when the entity is to receive information about the process or effort by the entity. It gives the entity a comparison to go on, but no indication of what the entity should do next time in order to reach better performances. This kind of feedback indicates that the entity's ability is what is of importance, and not the strategic work done by the entity. This is the reason for why this kind of feedback is not seen to be a good choice. Self-referenced feedback should be provided on unsuccessful achievements in order to provide the entity a picture of the current progress, and not about how far the entity is from reaching a potential goal. This kind of feedback provides the entity with an indication of how well the processes or methods that the entity has used, are [49]. **Examples of good feedback comparisons:** is when providing encouragement to reluctant entities, proven to be improving, despite the fact that the work done by the entity still not is good enough. But also when providing concrete comparisons between what the entity has delivered and what is expected. **Examples of bad feedback comparisons:** is when providing feedback to each entity based on different criteria or even no criteria at all, or by providing feedback, showing the entity's work right next to another entity's work.

The function of the feedback can either be descriptive or judgmental. It is common for entities to dismiss descriptive feedback when judgments are included, such as evaluative comments. In order to avoid this, to make sure that entities comprehend feedback as descriptive, entities should be allowed to practice and receive feedback without any kind of involvement of points or similar. This way, the entity eventually will learn that this is beneficial for them. Beneficial in what sense depends on the context in which this occurs. Another way of making sure that entities interpret feedback as descriptive, is to include a description of the potential errors, how close the entity is from the goal and what possibly could help the entity.

The valence of feedback can be positive or negative. Positive feedback is considered do be constructive criticism, meaning that positive feedback provides the entity with information about what has been done well and thus, what the strengths the entity possesses. Furthermore, the entity should be provided with points on improvement and guidance about what the entity can do to make these improvements. Guidance does not mean telling the entity exactly what needs to be done, but rather providing suggestions. Negative feedback on the other hand, should be accompanied with positive suggestions about what the entity can do to improve. In their book, *Teacher Feedback to Young Children in Formative Assessment: a typology*, Tunstall and Gisp suggest that feedback can be divided into two types: descriptive and evaluative [49].

When providing rewards, praise and such, it is considered to be positive evaluation feedback. Whilst, when providing feedback with punishments, criticism and similar, it is considered to be negative evaluation feedback. Descriptive feedback on the other hand, is always considered to be with a positive intention. Descriptive criticism is also considered as positive since the intent is for it to be constructive. Tunstall and Gisp further divide descriptive feedback into achievement feedback and improvement feedback. Achievement feedback tells the entity what in the work was well completed, whilst improvement feedback additionally provides advises of what strategies can be used and what improvements can be made [49].

Feedback clarity is of great importance in order to maximize the entity's understanding of the feedback provided. Proper vocabulary and concepts, known to the entity, should be used for the best possible comprehension. The amount and content of the feedback should match the entity's development level, and one should also make sure that the entity actually has understood the feedback provided. An example of bad feedback clarity, is when it consists of complicated sentences, too complex for the entity to understand. Or providing feedback about what oneself already knows, rather than giving feedback on what the entity in question knows and needs. And if one would simply assume that the entity understands the feedback without actually making sure that, that is the case, the feedback could be in vain.

Feedback can furthermore take different levels of specificity, which is: nitpicky, appropriate and general. The purpose of feedback specificity is to provide guidance, without actually completing the work for the entity. It is also intended to provide the entity with enough specified feedback so that the entity knows which steps to take further ahead, without completing the task for the entity. The degree of specificity should be tailored to match the entity and the task at hand. Errors or different types of errors should be identified and included in the feedback, without correction of each error. Providing good feedback specificity includes descriptions of concepts or criteria and potential learning strategies that could come in handy for the entity, in conjunction with the use of descriptive adjectives. Bad feedback specificity on the other hand includes the use of too many pronouns, vague suggestions and corrections of each and every error [49].

The final feedback content is the tone of feedback, setting the quality of the feedback, and can vary in implications and may affect what the entity will "hear". Conveying the tone of comments within the feedback is done both by the choice of words and the style. Choosing the correct words can lead to inspiring the entity, whilst discouragement can be caused by the wrong choice of words. The choice of words should be respectful towards the entity and position the entity as an active learner. Words that communicate respect, both for the entity and the work completed, should be chosen. In order to set the way for a good tone and word choice when providing feedback to an entity, words that assume the entity to be an active learner should be used. Furthermore, it is important to share what causes question-marks and also ask questions to the entity. Whilst, if choosing wordings that sets a tone of voice that indicates that the entity is being lectured or explicitly telling the entity what

should be done instead of also taking the time to listen to the entity, the feedback will not be considered as good [49].

6.2 Why feedback?

According to John Hattie and Helen Timperley, in their article, *The Power of Feedback*, they state that "*Feedback is one of the most powerful influences on learning and achievement, but this impact can be either positive or negative*" [47]. In his blog post, *Five reasons why feedback may be the most important skill*, Bob Dignen provides five reasons for why feedback is such a powerful tool [1]. In the first reason, which is that **Feedback is there whenever**, the author states: "*Every time we speak or listen to another person, in our tone of voice, in the words we use, in the silences which we allow, we communicate feedback – how far we trust, how much we respect, the degree to which we love, like or even hate the person in front of us. We cannot not give feedback. If we think we're not doing it, we're a dangerous communicator because it means we are probably not managing communication effectively*". This simply means that wherever we go, whenever that is and whomever we may communicate with, feedback will be involved.

Another reason for why feedback is one of the most important communication skills is due to that **Feedback is another way of showing effective listening**. This means that when one entity speaks to another entity, the former one wants to experience two things: the feeling of and knowing that they have been understood, and the feeling of that what was said actually was of value. If either of these two would to be removed, the entity speaking would quickly come to a sense of confusion or irritation. Therefore, when providing both aspects it means providing effective feedback.

Feedback is furthermore a good tool for motivating entities. When providing feedback, it is in a sense as providing praise and showing appreciation to an entity that has completed a job well, hoping to inspire the entity for completing the same job even better. Or even for completing future jobs in a good manner. This is a way of providing a greater positive feeling and commitment to those entities completing a good job. Just as feedback is a good way of motivating entities, it is also a great tool for **developing entities performances**. Feedback can normally be interpreted as criticism, but what it actually is, is constructive criticism. Its purpose is to work with under-performing entities in a constructive manner, developing their performances to a higher level. In these cases it is also of importance to use proper language. For example, it is not suitable to start a sentence by saying "You didn't do...", since it most likely will be interpreted as criticism. In such a case, it would be more appropriate to say something like "If you would have done this, it would have led to that instead" [1], properly explaining, in a constructive way, what should have been done instead for reaching a certain outcome.

Last but not least, **Feedback provides potential for learning**. When working internationally, especially, or dealing with customers with different nationalities than

oneself, there will be large business complexities and virtual teams, having problems with finding a common way of communicating, leading to events going wrong at times. In order to prevent such happenings, learning from ones mistakes, it is essential with feedback. The entity needs to invest time in asking the other entities about what their impression is of working with the entity. There will be times when entities will provide ill founded opinions about oneself, which might not be well received and tough to listen to. But it is after all an opinion and not a fact. As Dignen puts it '*with greater mutual understanding comes greater speed to market*". So, even though an opinion might not be accepted by the entity, it is still important that the entity manages to explain the reasons for saying and doing things a certain way. Basically justifying ones own behaviour [1].

7

Examples of Crowdsourced Information Retrieval Systems

There are today numerous applications and systems applying different methodologies in order to retrieve information through crowdsourced information retrieval. Also, the information retrieved is used differently in all systems. In the following section, applications adapting crowdsourced information retrieval will be presented. In the section thereafter, examples of organizations who have integrated trust management systems will be provided.

7.1 Crowdsourcing applications

Using crowdsourcing for retrieving information and content for applications is applied by numerous of already existing applications. One of them is **Sportswik**, which is an application whose content and data is retrieved entirely through crowdsourcing. In Sportswik, any kind of sports team can create their team profile and report the type of data they want, e.g. stream games, upload images from training camps or simply report results from games [50]. Once a team profile has been created, anyone in the community around the team can report content [51].

The way Sportswik works today, there is no kind of validation of the information provided by the crowd. The organization completely relies on their users to report the correct result, and thus hopes that mistakes will be noticed by other users, who in turn report it to Sportswik. This is a relatively hard task with the business idea that Sportswik has, since the organizations' main target group is the smaller, local football teams around the community, and for teams such as these, there is not as much coverage as for the larger, well-known teams. With low coverage, it means that it is hard for Sportswik to have any continuous validation of the information provided from users, and there is no option available for them to compare the received information with. Thus, there is no way today to both receive information and validate it for smaller football teams, other than relying on the crowd. Sportswik has thus chosen to crowdsource the entire application content and rely on the fact that users report correct information. Only time will tell, whether or not this will work in the long run.

Crowdscores, established in 2012, is another application whose content and data is provided to users via crowdsourcing. The application's main purpose is to de-

liver fast and correct match events, e.g. yellow cards, goals, fouls. This is done by allowing users to report the content to them [52]. An interview was conducted with a previous worker at Crowdscores, in order to retrieve information regarding how they handle the information retrieved through crowdsourcing. The information retrieved is partly handled by automation within a trust management system. This system retrieves the information together with that user or users trust values and depending on if the internal calculations seem to be correct in comparison with what the system is expecting, the event is accepted and displayed within the application.

At the early stages of the application, each user was assigned a trust value depending on previous performances. For an event to be approved and displayed in the application, a user's trust value had to be larger than a certain threshold that had been assigned to that specific event. The threshold depended on how popular the league and team was. As a precaution, Crowdscores always compared several users reports with each other, before displaying it in the application. This means that the organization compared the reports between each other and if they would prove to be the same and if the sum of these users' trust values was larger than the given threshold, the event would be approved and displayed within the application.

GoalShouter is yet another application that crowdsources its content, established in 2013. As the organization put it themselves "*GoalShouter is a free platform to create an amazing, professional live coverage for all matches of your soccer team! Create rich content for your supporters, bringing them with you always!*" [53]. The application allows its users to report anything that happens during a game and by doing so, create a forum opening up the possibility for other users to commentate and share the events and matches. It allows its user to generate anything from professional reports to statistical sheets, graphic content and videos. To begin with, the users need to create the match by themselves. The user that has created the match is referred to as a reporter. Once the match has been created, the reporter receives several options of events that potentially can take place within a match. When the reporter wants to report a certain event, for example a goal, the user selects the goal event, the team that scored and finally, which player within the team that scored the goal, and the event is automatically created and approved. The reporter furthermore has the option of taking snapshots of whatever event taking place on the field, and share it with its peers. Each reporter of the application can automatically share updated events on Facebook. Additionally, users of the application have the possibility of following each other. GoalShouter uses the term fans, for those users that follow a reporter. These fans have the possibility of commentating and sharing the match, that the reporter has created. The fans can also vote on question that the reporter asks, such as "Who was the best player in the game?" [54].

Another feature of the application, creating an incentive for the reporters to continue providing information, is that the direct feed of a match can be directly embedded into the reporter's own website. This increases the reporter's and its sponsors visibility, making the reporter's team known to a large amount of people all over the world. This becomes particularly useful, if the team that is being reported, is a

smaller team with otherwise low coverage and low fan base [54].

ScoreStream is a high-school sports application completely driven by the fans. As the organization put it themselves "*We aim to empower fans, schools, teams, and media companies with the tools to provide accurate, real-time sports coverage.*" [55]. All match schedules, scores, photos and videos are all provided by the fans of the different teams within the application, meaning that all the content within the application is retrieved through crowdsourcing [56]. The fans can follow their favorite teams within the application, where the coverage is provided by the fans themselves. What the fans get from the application, besides the possibility of supporting their teams, is real-time scores and data in general. The highlights of the games can be instantly shared on social media. The photos that the fan takes and wants to add on the team's page, can be edited with filters and overlays directly within the application. The fans can furthermore live chat with other fans and thus cheer on their team together, creating a community for the fans [55].

Besides the fans, the application is directed both towards media and the teams and schools. For the media, the data coverage provided by the users, drives the partner's of ScoreStream, by providing real-time scoreboards, photo and video galleries. Since the tool is free, it provides a good head start in order to be the leader in ones own local sports. All the real-time scores, images and videos on the website are free of charge. This provides a platform for the media to engage the fans to use the tool, leading to greater experience for the fans, which in turn will lead to better content [55].

For the teams and schools, the platform becomes a major brand experience, according to ScoreStream themselves. The real-time scores, images, videos and alerts provided by the fans can be instantaneously shared from the same platform, providing the team with all tools required in order to share the most important moments. A team member can become a ScoreStream General Manager, meaning that this member can customize the team's page and unlock onlin tools, powerful for the team's display [55].

Wikipedia is a free encyclopedia, driven by the non-profit, charitable organization called Wikimedia Foundation. According to the Wikimedia Foundation, their goal is to "encouraging the growth, development and distribution of free, multilingual, educational content, and to providing the full content of these wiki-based projects to the public free of charge". Furthermore, with this, they want to imagine a platform where users freely can share knowledge with each other [57]. All projects are developed collaboratively by users and thus operates under the so called free Creative Commons License [58]. This license enables anyone to freely use, edit and copy the content within each project. This way, of having the content being completely generated by users, has been the rise of Wikipedia. However, this openness has also given Wikipedia numerous challenges to tackle since the trustworthiness of the content becomes hard to decide for users visiting the site. This is mainly due to that articles in Wikipedia are under constant change, where its authors vary from domain experts to the opposite. Further-more, users visiting the site do not receive

any information regarding an articles state prior to the changes made giving the latest version. Wikipedia is evidently a platform in a great need to make appropriate changes in order to establish trust with its users and provide evidence of that the content is trustworthy, or at least an indication to what extent the content can be trusted.

7.2 Trust management systems

EBay is a typical, and one of the most well-known, examples of a system using reputation-based trust management, which enables its users to rate each other based on their interactions. Leaving feedback can only happen after an actual action has taken place between two users, where both the sellers and buyers have the possibility to rate each other. The seller may only give the buyer positive or no feedback at all, whilst the buyer can leave negative, neutral or positive feedback about their experience with the seller [59]. The reputation score for a user is calculated by taking the positive feedback minus the negative feedback. This reputation score is then displayed on the users' different auction pages and is available at all times for other users to see. There is also the possibility of seeing a more detailed description of each feedback type; positive, negative and neutral, across different time spans [60]. This is one of the most valuable aspects in EBay's system, since it gives the users a possibility to build up a reputation and thus get the possibility to sell and buy even more in the future.

TripAdvisor is yet another example of a reputation based trust management system. TripAdvisor allows its visitors to rate their stay at a hotel, both by a bubble rate between one to five bubbles, where one bubble means "terrible" and five means "excellent", and by writing a comment. The overall rating made by users for one specific hotel is a measure of the quality based on users' experiences with the hotel, consisting of all of the user's ratings added together, forming an average rating for that hotel. The overall rating takes numerous of values into its decision making, such as: the quality, quantity and how recent the review is. The quality of a review is represented by the five-bubble rating, giving a hotel with five bubbles a better quality than one with four. In the algorithm, TripAdvisor also takes into consideration how recent the reviews have been given, meaning that reviews that have been received closer to present time are higher ranked than older ones, and thus get a higher value. The reason for this is due to that users care about the hotel's present state, and not how it was a couple of years ago, which thus makes older reviews less interesting. Lastly, if the quantity of reviews for a certain hotel are more than they are for another hotel, it does not mean that the former one will be ranked higher. Rather, TripAdvisor's algorithm is only interested in having enough reviews for a meaningful comparison statistically wise [61]. The overall rating can furthermore be broken down into the distribution of ratings, what kind of traveler it is or how good the business is [62]. Except for the overall rating, the individual rating is also displayed next to each person's review [63].

8

Design of the Trust Management System

The trust management system developed for the purpose of this study will have two main tasks; calculate the trust value based on users previously reported lineups, and eventually make automated decisions based on the trust value, regarding whether or not to approve a users lineup. The concrete requirements for the trust management system are presented in Section 3.2. These requirements were gathered during discussions with co-workers at Forza Football and during a brainstorming session, which is presented in Section 8.2. The gathering of these requirements is a part of the Relevance cycle presented in Figure 1.1, and has the purpose of gathering the business needs and requirements from the environment for which this trust management system is to be designed for. These requirements are to be met by developing the trust management system iteratively during four different iterations where, for each iteration, a new metric is added to the system and then evaluated through validating sessions with participants relevant for this study. Each of these iterations are presented in their corresponding section. Before the design and development of the trust management began, relevant knowledge for the study had to be gathered, in order to cover all the terms and most important aspects related to this study, which is presented in the following section. Alongside, before the first iteration began, a brainstorming session was conducted with a co-worker at Forza Football AB, which is presented in the section thereafter.

8.1 Initial knowledge base

As of the current situation in Forza Reporter, there is a need for automated decision making when receiving reported lineups from users of the application. The main issue existing is knowing when a user can be trusted and the information provided is trustworthy enough to be used. The trust of a user and how it is managed, are the main components in the knowledge base for this study and is the information that mainly has been gathered during the Rigor cycle, as shown in Figure 1.1. Besides this, the primary knowledge base further consists of the following components, setting the base for this study: Trust Management Systems, Crowdsourced Information Retrieval, Trust Management Systems in Crowdsourcing and the Process of Feedback. Furthermore, examples of systems adapting crowdsourcing and trust management system has also been provided.

The knowledge has mostly been based on secondary data gathered from the web. This due to the given wide range of information existing on the web. The gathering of knowledge was primarily conducted by searching the web for key words, such as trust, trust value, trust management systems, crowdsourcing, crowdsourcing systems etc. By doing so, numerous books, articles and case studies were found, relevant for the study. In order to expand the knowledge base even further and find even more sources for the information already gathered, the references used in each paper were also reviewed. By doing so, numerous, additionally good papers were found, useful for the study in many ways. For the first research question, **RQ1:** *Does using the trust value of a user to evaluate the trustworthiness of the information provided from that user yield similar results as a manual evaluation?*, an extensive information gathering was conducted. Entire Chapter chapter 3 creates a solid foundation for answering this research question. The different evaluations conducted on the trust management system created for this particular study, will however provide even more data points for answering this research question.

As for **RQ3:** *How should the user receive feedback regarding what should be done to reach a certain trust value?*, research was conducted on existing papers on feedback in order to find information about what good feedback is and how it is given in the best possible manner. Some of the papers found, were directed towards feedback to students, co-workers etc. Despite this, the papers were still found to be applicable in other contexts, since the same feedback can be applied to any crowd. Furthermore, the points brought up by the authors are relevant for the case of Forza Reporter and thus, this study. In order to make sure to give good and informative feedback, it is of great importance that the correct strategies and content is chosen, and entire Chapter 6 provides a solid background for this, providing information about what good feedback is and how it should be given. This information has provided further steps into answering research question three.

8.2 Brainstorming session with focus group

Before diving into details and the actual development of the trust management system, a brainstorming session was conducted with a co-worker at Forza Football. The session was conducted with a backend developer at the organization who has previous experiences in working with trust management systems, and thus providing insightful opinions and feedback about the different options of solutions for the problem at hand. The brainstorming session is mainly part of the relevance cycle, as shown in Figure 1.1, since it provides further requirements to the study and what the business needs are in the context of Forza Football.

This brainstorming session was semi-structured, allowing the participant to answer both shortly at questions, but also to provide potential valuable explanations, mainly from previous experiences that this participant has. Firstly, the session began with presenting the problem statement, as shown in Chapter 2, along with the requirements for the trust management system that had been set prior to this state, as presented in Section 2.2. With this as a basis, it opened up for free discussions

regarding how to solve the problem at hand in the best possible way. This discussion was initiated by asking the participant what previous experiences relevant for this study the participant has. The participant was asked to, if possible, provide information about how the issues that this study has been derived from, had been tackled at the participant's previous work place. By having some insights in how a trust management system can be integrated into a crowdsourcing application, the discussion continued on to what would be the most suitable design of a trust management system for the case of Forza Reporter. The discussion mainly revolved around what the initial trust value should be and what the increase and decrease of the trust value should be when reporting a correct and incorrect lineup, respectively. The discussion furthermore continued on to how the trust value should be used in the trust management system, whether it should be a global trust value or a local one, meaning if the trust value only should be valid for a certain team or league, or if the same trust value should be valid throughout the entire application. The results from this brainstorming session are presented in Section 9.1.

8.3 Iteration 1

After the brainstorming session, a spreadsheet was created consisting of all users reported lineups from when Forza Reporter first was released until the day the sheet was created. The reason for creating such a sheet, is both due to internal needs at Forza Football, but also since this sheet provides all information available about the users taking part in providing lineups in Forza Reporter. It is information such as the number of reported lineups for each user, when the lineups were reported and the approval and rejection rate, which will in particular be of relevance for the third iteration, presented in Section 8.5. Since the number of reports until this point were a couple of thousand, and there is no need in showing the entire document for the purpose of this study, only a short example was added into this report, and can be found in Appendix Appendix A, Figure A.1. Within the sheet, the user-id, the time of when the report was sent to Forza Reporter and an indication of if the reported lineup was approved or discarded, is provided.
In this first iteration, an initial version of the trust management system was designed and developed, with a function of all reported lineups for each user, in order to retrieve the trust value for each user. The discussions that took place during the brainstorming session, presented in Section 8.2, gave a solid foundation for where the first iteration should begin and what it should include. Furthermore, the requirements that in particular were used for this iteration are:

- Assign trust values to users based on previous performances. The value lies between 0-1. 0 indicates that all the user's reported lineups have been rejected. Having 1 in trust value indicates that the user's Z most recent lineups have been accurate, meaning that the user is seen as completely trusted. How many lineups, Z, need to be reported will be tried out during the different iterations in this study.
- When reporting the first correct lineup, what trust value should the user receive?

- What happens when the user reports an incorrect lineup?
- Should the user report x correct lineups in order to get a higher trust value? Or does the trust value increase with a certain value X for each correctly reported lineup?
- Increase the users trust value if the reported lineup is correct, even though the lineup might not be accepted.
- Decrease the users trust value with Y when an incorrect lineup is reported.
- Should the trust value be global or local? Meaning, should the trustworthiness of a user be valid throughout the entire content in the app, or should it only be valid for the specific team that the user has reported a lineup for?

As a part of this iteration, validation sessions were conducted in order to assess the significance of the trust value, and what a human might make of it. These validation sessions are a part of the design cycle, as presented in Figure 1.1, where the trust management system is designed and then evaluated through validation sessions with participants relevant for this study. The theory was evaluated internally by having co-workers at Forza Football AB making independent decisions about whether or not to approve a user's lineup depending on two distinct criteria: the trust value of a user and by viewing that user's history of reported lineups. Worth noting is that the same person did not get to validate both criteria, but rather only one of the criteria, in order to not influence the decision making of the person in any matter, and in order to be able to make proper comparisons between the results. By independent decisions it is referred to that the participants in the validation sessions were not affected by any external forces and they could not communicate with each other in any sense. In order to make sure that this did not happen, the participants got to complete the validation at different times, starting with the participant who got to make decisions based on the second criteria. Worth noting is that the participants completed the validation session individually, no matter if validating the same criteria, and not in a group. The reason for conducting the validations using these two criteria, was in order to receive an indication of the significance of the trust value and if it gives the same outcome as viewing users histories of reported lineups. This validation session has also provided a further step into answering **RQ1:** *Can trust be used in order to distinguish sources of trustworthy information from others?* and **RQ2:** *How should the trust value be used once assigned to the user?*.

8.3.1 Participants for validation sessions

As presented in Section 1.3, in Figure 1.1, the environment for which the trust management system is to be designed and evaluated for is the crowdsoured information retrieval domain, which consists of the people in it and the problems which this study has the aim at solving. As for the people in this domain, they consist both of the users of the applications that the organization Forza Football has developed, but also of the creators of these applications. The users should not take part in any internal processes, but rather use the services provided by the organization. Furthermore, it is the creators of the applications who are responsible for the internal processes, how

they are designed, evaluated and used, and who thus know it the best. Therefore, since the trust management system is an algorithm processed internally, meaning that it is not of significance for the users, but is rather only used by the creators of the application in order to enable automated decisions, the users of the application have not been a part of the validation sessions taking place in this study. In order for the results from this study to be as accurate as possible, and in order for the results to reflect reality, the participants for the validation sessions were carefully chosen, only including co-workers at Forza Football who either have worked in the administrative user interface or with direct development of Forza Reporter, and thus possess the relevant knowledge, making it a total of six participants. Furthermore, since quality is of a larger significance for this study than quantity, the number of participants is enough. The same number of participants and the same individuals will take part in all validations sessions presented hereinafter.

8.3.2 Validation session based on users history of reported lineups

This validation session started off by letting each participant know what the validation session was going to be about and what was to be expected. The participants were asked to, based on seeing a user's history of reported lineups, in conjunction with the decision that was made regarding each lineup, make a decision of whether or not to approve the next lineup reported by the user. The participant was provided with printouts, as Figure A.5 in Appendix A exemplifies. Each history of reported lineups corresponds to an aggregated trust value, unknown to the participants evaluating the criteria of users' histories. The validation was conducted using real users data currently existing in Forza Reporter, with history that corresponds to each trust value existing in the spectrum of 0-1. The reason for doing it this way, is due to that it in particular is worthwhile when making comparisons to the results from the second validation session in this iteration, explained in the next section. These participants did not however receive any information about the trust value, or that it even exists, since the participants were not to be affected in any manner, but was rather expected to focus only on making decisions based the user's history of lineups.

Once each participant had been provided with relevant and enough information to complete the validation session, the participants were shown printouts of the history of a users reported lineups, for a timespan from December 2017 until March 2018. The order in which the different printouts were displayed, was randomly selected starting with the user's history which corresponds to a trust value of 0.9. This same process was continuously completed with users histories, corresponding to the other trust values existing in the spectrum as well. The order in which the printouts of users histories were shown, corresponds to the following trust value: 0.9, 0.2, 0.5, 0, 0.7, 0.1, 0.4, 0.6, 1.0, 0.3, 0.8. During the validation sessions, the participants were also asked to think-aloud regarding any thoughts that might come to the their mind. The participants were asked to furthermore explaining the reasons for making a decision of accepting or discarding a lineup based on users' histories of reports. No interference was made during this process, and notes were taken on all

opinions expressed by the participants. Worth noting is that the participants did not complete the validations in a group, but rather got to complete the validation at different times, individually, as presented in Section 8.3. The results from these validation sessions are provided in Section 9.2.1.

8.3.3 Validation session based on users trust values

This validation session started off in the same manner as the previous one, with letting the participants know what this validation session was going to regard. The participants were asked to, based on users trust values, make a decision regarding whether or not to accept each users lineup. The participants were provided with screen-shots resembling the administrative user interface, that had been printed out. The design of the screen-shots was made to match the real administrative user interface used for Forza Reporter, in order to keep the situation as similar as possible to reality. An example of the printouts is shown in Appendix A, Figure B.10. Since the order of the trust values may affect the participant's decision making, the trust values were shown in a random order, as follows: 0.9, 0.2, 0.5, 0, 0.7, 0.1, 0.4, 0.6, 1.0, 0.3, 0.8. Before showing the printouts to the each participant in the given order, the participants were provided with relevant information. The information provided to these participants was only about what the trust value is, meaning that the participants were told that the trust value is a numeric value between 0-1 used for distinguishing trustworthy users from those who are not. The participants were furthermore also told the two limits existing for the trust value, and what those two values mean: 0, meaning not trustworthy, and 1, meaning trustworthy.

Once the participants had been provided with relevant and enough information to complete the validation session, the participants were shown the different screen-shots of the administrative user interface, with the corresponding trust value for that user. The order in which the different screen-shots and trust values were shown, was randomly selected, as presented above. The same process was conducted for each trust value existing in the spectrum between 0-1. During the validation session, the participants were asked to think-aloud, sharing any thoughts, insights and question marks that might occur during the process, which all was recorded in a document simultaneously. The participants were asked to make a decision of whether or not to approve the lineup only based on the trust value, and provide a reason for why a certain decision was made. This same process was completed with all the other trust values existing over the given spectrum. Just as for the other validation session, no interference was made during this process, and notes were taken on all opinions expressed by the participants. Worth noting is that the participants did not complete the validations in a group, but rather got to complete the validation at different times, individually, as presented in Section 8.3. The results from this validation session are provided in Section 9.2.2.

8.4 Iteration 2

In this iteration, another value was added to the trust management system, which is reviewing the quality of the lineup reported by the user. As per one of the requirements, provided by Forza Football, for the system, as given in 2.2: "*Should the increase of the users trust value depend on the correctness of the lineup? There are cases where the lineup provided by a user is accepted, first after certain, minor adjustments have been made. That would mean that the increase of the users trust value would thus depend on the number of adjustments required prior to the acceptance of a lineup. Or should the trust value be increased with the same value no matter the case*?". This requirement is a rather important one, since users who make mistakes should not get the same increase in their trust value as a user that has reported a completely correct lineup, but since their lineup gets approved, first after some adjustments have been made, they should still receive some recognition and a slight increase in their trust value. In order to make conclusions about how many mistakes are acceptable, which mistakes are acceptable and which mistakes are seen to be worse than others, a shorter interview with open-ended questions was conducted with a co-worker at Forza Football.

8.4.1 Open-ended interview

This interview started off by providing the co-worker with information about the current status of the project, what had been done at this point and what the plan was for going forward. The co-worker was provided with the information, that this iteration was about the quality of the lineup and that the quality of a lineup is considered to be the number of adjustments required prior to the acceptance of a lineup. This led on to a discussion about which mistakes completed by users, first had been modified in the administrative user interface, and then approved at this point in time. This continued on into the main question of this interview, which was: "*Which of the adjustments allowed at this point in time, are worse than the others?*". The bigger parts of the interview were spent on this question due to different viewpoints on the matter. But after exemplifying the matter, providing both advantages and disadvantages with different approaches, conclusions were made. The results from this interview are presented in Section 9.3.

8.4.2 Validation sessions

As a part of this iteration, validation sessions took place in order to assess how the added metric of quality of lineups might affect the decision a person makes regarding approving or discarding lineups. These validation sessions are a part of the design cycle, as shown in Figure 1.1, where the trust management system continuously is designed and then evaluated through validation sessions with participants relevant for this study. The evaluation was conducted internally by having the co-workers at Forza Football AB making independent decisions about whether or not to accept a user's lineup depending on two distinct criteria: the trust value of a user and by

viewing that user's history of reported lineups. The methodology was the same, as described in Sections 8.3.2 and 8.3.3. The only difference in these validations was that another value had been added into the calculation of the trust value. For the participants making decisions based on users history, they got to see another value, which was the quality of each lineup, meaning if the lineup is completely correct, correct with one, two or three adjustments, or simply incorrect. The results from the validations conducted in this iteration are presented in Sections 9.3.1 and 9.3.2.

8.5 Iteration 3

Based on the results retrieved from the first two iterations, which are presented in 9.2 and 9.3, this iterations purpose will be to add more metrics, alongside the trust value, in order to be able to make conclusions regarding if the trust value alone is significant enough in order to make an automated decision regarding the information provided by users, or if any of the added metrics are required to be used in conjunction with the trust value. Another possible outcome from this, could be that one of the added metrics is more useful than the trust value. In this iteration, users histories of reported lineups is no longer valuable to use for validation purposes, meaning that this iteration only will use screen-shots from the administrative user interface. Each metric that could be of value for making a decision regarding users lineups, has been added next to the trust value in the administrative user interface. The values that were evaluated, with the trust value, are: users approval rate, users trend, which is the number of approved lineups out of the five last reported ones, and users approval rate together with the number of reported lineups. In order to be able to make conclusions regarding the significance of the trust value, a validation session will also be conducted using users' approval rate and number of reported lineups, only. Meaning that the trust value is completely excluded in this validation session. Besides the aim of providing further data for the requirements as specified in Section 8.3, this iteration furthermore strives at handling the following requirements.

- Should the system only rely on the user's trust value or by comparing users lineups for the same team with each other, when making automated decisions? Or both? Or does the automated decision making process need to be supported by another value, indicating the user's history of reported lineups in some manner, for example?
- Should a user get its lineup automatically approved first when its trust value is above a certain given threshold?
- If setting a threshold for what trust value is required for getting a lineup approved, what happens with lineups reported by users, if their trust values are below the threshold, but their lineups have proven to be correct after comparisons between each other?
- Should a user with a trust value beneath a certain threshold automatically get its lineup discarded? And leave room for it to be manually reviewed? Or should the user's lineup automatically be compared to other users lineup for the same team, and let the system make a decision based on their similarities?

- If comparing users lineups with each other in order to decide whether or not the lineup is correct, should the users' trust values decide who gets its lineup displayed in Forza Football, or should that be based on which user reported the lineup first?

The validation sessions conducted in this iteration are also a part of the design cycle, as shown in Figure 1.1, where the trust management system is designed and then evaluated through validation sessions with participants relevant for this study. As for these validation sessions, in order to gather as many data points as possible, all participants that took part in the validation sessions conducted in the previous iterations, are all part of the validation sessions for this iteration as well. This means that there were six participants in total for this iteration. Three of these participants, participant, have prior knowledge about the trust value, since they took part in the two previous validation sessions where the trust value was evaluated, only. Thus, the only information required to be provided to these participants, was the meaning of the other metrics used alongside the trust value, which are rather self-explanatory. As for the other three participants, who in the previous iterations got to make decisions based on users' history, only, had to first be given information about the trust value, before receiving information about the other metrics. The information provided to these participants was the same as described in Section 8.3.3. Once all the participants had been provided the proper information for completing these validation sessions, they were shown screen-shots from the administrative user interface, with the various metrics. All validation sessions were conducted in the same order, starting with the one where the trust value was evaluated together with users' approval rate, followed by an evaluation of the trust value together with users' approval rate and the number of approved lineups, and ending with a validation session where the trust value was evaluated together with the number of approved lineups out of the five last reported ones. These first three validation sessions all evaluated the trust value alongside another metric, whilst the final validation session for this iteration evaluated the number of reported lineups together with users' approval rate, completely leaving out the trust value. In the end of the last validation session conducted in this iteration, each participant was asked the following questions: "*Which of the metrics made the decision the easiest?*" and "*Do you think that this metric is enough for making a decision regarding a lineup, or would you still like to have the trust value?*". The results from these validation sessions are presented in Section 9.4.

8.6 Iteration 4

Feedback is of great importance, because it shows users that the information they provide actually is being used, or at least acknowledged. Furthermore, feedback is a key for giving users an incentive to do more and better, and the feeling of recognition is what many seek [1]. Since the current state of Forza Reporter lacks any kind of feedback, there is evidently a great need to integrate feedback into the current process, in order to provide incentives for users to report more and better lineups. Therefore, this last iteration for this study has the purpose of validating what kind

of feedback users should receive, and especially, if the trust value should be included in the feedback to users, or if it simply is insignificant to them.

The requirements that in particular were used for this iteration are the following:

- Should the trust value be displayed to the user? Will the user understand what that number means? Or is this not significant to the user?
- Should the trust value be gamified? Meaning, if a user for example has a trust value of 0.1, the user is a water-boy. Or if the user has 1.0 in trust value, the user is a coach.
- Should the trust value be displayed together with its gamification to the user? Or only the gamified value?
- Should the user receive feedback about what is required to reach the next step or even the highest value?
- When the lineup is correct or the user's trust value is high enough to get the lineup automatically accepted, send proper feedback explaining the reasons for the acceptance.
- When the lineup is incorrect, but the user has a high trust value, send feedback explaining the reason for rejection.
- When the lineup is incorrect and the user's trust value is low, send feedback explaining the reason for rejection.
- For each rejected lineup, should the feedback include the new trust value, the reason for it being lowered, and an explanation for what the user has to do in order to reach a higher trust value again, or only the reason for rejection?
- For each accepted lineup, should the feedback include the new trust value and an explanation about what that new value means, or only the reason for acceptance?
- What proper feedback is, will be investigated during the different iterations in this study.

For the validation sessions in this iteration, all six participants were once again involved. The participants were shown two different drafts of user profiles; one including the trust value and one including a gamification element. Both drafts contain some other data points as well, that could be relevant for the user, but also in order to keep the situation as similar to reality as possible. The drafts used for these validation sessions are presented in Appendix D, Section D.1. In conjunction to this, an open-ended interview was conducted. The questions asked in these interviews are presented in Appendix D, Section D.2.

9
Results

The design science research methodology used throughout the study was completed in different iterations, where the insights from a previous iteration was additional knowledge for the iteration thereafter, where further metrics and complexity was added to the trust management system in order to retrieve more data for answering the research questions, presented in Section 1.1. In order to get as good results as possible from the validation sessions, relevant participants had to be chosen. The selection was carefully made by selecting co-workers at Forza Football AB who in some manner have been involved in Forza Reporter, may it be the direct development of the product or the work in the administrative user interface, with approving or discarding lineups, as explained in Section 8.3.1. This gave a total of six co-workers taking part in all of the validation sessions. Since there are six participants, the participants will be referred to as *Participant x* throughout the coming sections. X will be between A-F and refers to each participant. If referring to participant A, for example, throughout several sections, it is the same participant being referred to.

Within this chapter, the results from the brainstorming session and each validation sessions in the various iterations will be provided. Potential insights and interpretations of the results will be provided in the chapter thereafter.

9.1 Results from brainstorming session

When providing the participant in the brainstorming session with the requirements for this study, the open discussion that began thereafter, provided room for insightful feedback and conclusions for how to best tackle the issues at hand.

From Section 2.2, it is evident that the users are to receive a trust value based on their previous performances. This means that a user with a record of providing correct lineups will receive a higher trust value in comparison to a user who has a record of bad reports or even a user who has a record of both inconsistency. The trust value given to users is a numerical value between 0-1. A user who continuously has provided correct lineups, will eventually reach a trust value of 1, indicating that this user is trustworthy. A user who has proven to not be trustworthy, on the other hand, by reporting bad lineups, will eventually reach a trust value of 0, indicating that this users is not trustworthy, and therefore neither the information provided by this user. A user who however is new to the application, has neither proven to be trustworthy nor the opposite, meaning that this user should not start on the

same trust value as a user who has proven to not be trustworthy. On the other hand, these users have not proven to be trustworthy either and should therefore not receive a too large trust value as a starting point. Since the next trust value after 0 is 0.1, it is appropriate that new users receive this trust value as a starting point. For each correctly reported lineup, users should receive an increase to their trust value. During the brainstorming session it was decided that the trust value should increase with 0.1 for each correctly reported lineup. The reason behind this is, if the trust value starts at 0.1 and a user receives an increase to its trust value with 0.1 for each correct lineup, that means that the user needs to provide nine correct lineups in a row in order to reach a trust value of 1, and thus be seen as completely trustworthy. Providing nine correct lineups in a row is enough data points to evaluate a user's performances, since the user has shown consistency, and thus, trustworthiness. For a user proven to not be trusted, having a trust value of 0, it would require ten correctly reported lineups to reach a trust value of 1. This increase in the trust value should happen for all correct lineups, even though the lineup might not have been reported fast enough to be displayed in Forza Football. This due to that there can be several users reporting a correct lineup, but only the fastest one will be selected for being displayed within Forza Football. And no matter it being approved or not, if the lineup is correct, the user should get recognition for it. In Section 2.2 it was mentioned that there are cases where users report a lineup where a couple of adjustments are required before it can be approved. One such adjustment, is incorrect players, where the mistakes made by the user are caused by the fact that some players required for the lineup do not exist within the application yet. Since this kind of mistake in essence is not a conscious one by the user, the user should receive an increase of 0.1 to the trust value, if the lineup ends up being approved. On the other hand, there are users who make minor mistakes, which first after some adjustments end up being approved. Mistakes such as these were mentioned in Section 2.2, and are typically a switch in 2-3 players positions. Since these lineups end up being approved, first after a few adjustments, the users should still receive an increase in their trust value, but not an as large increase as a user who reports a completely correct lineup, not requiring any adjustments. There are however users who make obvious and conscious mistakes, leading to the lineup being discarded. Making such a mistake should have larger consequences than a correctly reported lineup should be awarded. A discarded lineup should therefore receive a decrease of 0.2 in the trust value. Note that a user can never reach a trust value lower than 0 or higher than 1.

During this brainstorming session it was also discussed whether the trust value associated with a user should be global or local. That a trust value is global means that the user will be as trustworthy for team X as for team Y. If the trust value is local however, it means that the user will only be seen as trustworthy for team X. If the same user would start reporting for another team, Y, it would mean that the user receives a completely new trust value, indicating the trustworthiness in respect to team Y. Since the majority of the users reporting lineups so far have been reporting on teams that they are familiar with, that they play in themselves or that they even coach, and thus hold valuable information, there is no need to use a local

trust value at this point. It has thus been decided that the trust value will be global.

9.2 Iteration 1

For this iteration, an initial version of the trust management system was developed, based on the requirements presented in Section 8.3. The version consists of a function providing each user a trust value, based on their history of reported lineups. Each users initial trust value starts at 0.1. The function then takes, for each user, all reported lineups from the start until the last reported lineup by that user, and for each lineup checks whether it was correct or incorrect. For a correct lineup, the trust value increases by 0.1, whilst 0.2 is withdrawn for an incorrect one. After scanning through all reported lineups for a user, the function provides the user's current trust value. Once the function has computed, the user's trust value will be provided. This trust value is saved for future references, so the next time the user reports a lineup, the trust value will either be increased by 0.1 or decreased by 0.2, providing the user's new trust value. The user's trust value is then displayed within the administrative user interface, meaning that the ones working at Forza Football, currently responsible for approving or discarding lineups, will be able to view each users trust value.

Once the implementation of the initial version of the trust management system had been completed, validation sessions took place with co-workers at Forza Football, which is presented in the next two sections.

9.2.1 Validation session based on users history of reported lineups

This validation session was divided into three different sessions, with three different co-workers participating in one session each. This way, three different aspects were gathered, without allowing the participants to be affected by any external forces or by each other. All three sessions were conducted in the same manner and in the same order, to keep it consistent. In Section 8.3.2, it was mentioned that the test session started off by giving the participants relevant information, following with showing screen-shots of users history of reported lineups, corresponding to an aggregated trust value. The results from this test session are presented in Table 9.1. Note that the results of the examples shown of a users history of reported lineups, corresponding to a certain trust value, are presented in an increased order of trust values, and not in the order in which they were shown to the participants in the validation sessions. The first column refers to each history, corresponding to a certain trust value, that was used for the validation session. The other two columns show the two possible decisions that the participant could make, which was to either approve or discard a users lineup based on the history. For each participant, the decision and the reason for the decision are presented in the table.

The history of a user as displayed in Appendix A, figure:	Approve	Discard
A.2 (Corresponding to a trust value of 0)		**Participant A:** Discard the lineup based on the recent reports being discarded. - - - - - - - - - - - - - **Participant B:** Would like to double check the actual lineup, if possible. But in general, there are too few data points. - - - - - - - - - - - - - **Participant C:** Too few data points, with the last two lineups being discarded.
A.3 (Corresponding to a trust value of 0.1)		**Participant A:** A lot of inconsistency in the user's history, with many discards. Approved lineups seems to be a lucky run. - - - - - - - - - - - - - **Participant B:** Very few data points, with than half discarded. Would double check the lineup, but most likely discard it. - - - - - - - - - - - - - **Participant C:** More discarded than approved, and too few data points.

Table 9.1: Results from first validation session, based on users histories.

Continuation of table 9.1

A.4 (Corresponding to a trust value of 0.2)		**Participant A:** Discard the user's lineup due to that there is not enough data - one discard, one approved. There is no evidence of consistency. - - - - - - - - - - - - - **Participant B:** Similar reasoning as A. Would not auto approve it. - - - - - - - - - - - - - **Participant C:** Too little data. Would not approve automatically.
A.5 (Corresponding to a trust value of 0.3)		**Participant A:** Large inconsistency, with a long list of discards, especially on the same day. There is however a good change that then next lineup could be correct, but due to the inconsistency, discard the lineup. - - - - - - - - - - - - - **Participant B:** Would like to double check the lineup. Feels like a user who sends in preferred lineups, and not the official lineup. - - - - - - - - - - - - - **Participant C:** Many reported lineups, but more discarded than approved. Improves over time, but not too good to be approved automatically. Reports often = power user. But would not manage without administration.

Continuation of table 9.1

A.6 (Corresponding to a trust value of 0.4)	**Participant A:** Improvement in the most recent lineups, after providing really bad ones. Bases decision on the five past approved lineups in a row. - - - - - - - - - - - - - **Participant B:** Similar reasoning as A.	**Participant C:** Many lineups over a short period of time. Almost 50/50 as for approved and discarded. Would not automatically approve.
A.7 (Corresponding to a trust value of 0.5)	**Participant A:** Good recent track record, with several approved lineups on the same day. Discarded lineups does not seem like a normal behaviour. - - - - - - - - - - - - - **Participant B:** The user used the app during two different days, with almost everything approved. Would approve since last five are approved.	**Participant C:** 20% of the lineups are incorrect. Would thus not approve automatically.
A.8 (Corresponding to a trust value of 0.6)	**Participant A:** Based on a good track record of approved lineups, and since the user only has approved lineups, the next reported lineup by this user would be approved. - - - - - - - - - - - - - **Participant B:** Similar reasoning as A.	**Participant C:** Not enough lineups or data. Six lineups is too little to make a decision.

Continuation of table 9.1		
A.9 (Corresponding to a trust value of 0.7)	**Participant A:** A good trail of approved lineups, based on the last lineups. - - - - - - - - - - - - - **Participant B:** Seven lineups in a row approved. However, it is only during one hour, same day. Need more data points over longer time period.	**Participant C:** Only one discarded lineup, which also was the first one. Would however not automatically approve without double checking the actual lineup.
A.10 (Corresponding to a trust value of 0.8)	**Participant A:** The user seems trustworthy, consistent and regular. Discarded lineups is not common - no two lineups after each other where both have been discarded. - - - - - - - - - - - - - **Participant B:** Similar reasoning as A.	**Participant C:** Reject, due to a ratio of 20% of lineups being rejected.
A.11 (Corresponding to a trust value of 0.9)	**Participant A:** Many lineups reported, and the majority are approved. The user has only gotten one discarded lineup. - - - - - - - - - - - - - **Participant B:** Similar reasoning as A. - - - - - - - - - - - - - **Participant C:** The user has many reported lineups, with only one mistake. So would give this user a pass and approve the next lineup.	

Continuation of table 9.1		
A.12 (Corresponding to a trust value of 1.0)	**Participant A:** Only one discarded lineup. Reports are provided on a weekly basis, and the user is consistent. - - - - - - - - - - - - **Participant B:** Similar reasoning as A.	**Participant C:** One incorrect lineup, which was quite recently, giving a total ratio of 8% of discarded lineups. Meaning that the error is most likely not due to a misunderstanding from the user. Would not automatically approve.

9.2.2 Validation session based on users trust values

This validation session was divided into three different sessions, with three different co-workers participating in one session each, just as described in the previous Section, 9.2.1. As mentioned in Section 8.3.3, the test session started off, after giving each participant information about the test session, with showing a screen-shot of the administrative user interface with a user who has reported a lineup with a trust value of 0.9, as shown in Appendix B Figure B.10. The participants were expected to make a decision regarding whether or not to approve users lineup based on the trust value only. The first participant's, referred to as **Participant D**, initial reaction was not any obvious decision. The reasoning that the participant had was that, if this user has started at a trust value of 0 and then slowly improved its performances and thus reached a trust value of 0.9, the participant would approve this user's lineup. On the other hand, if the user previously has had a trust value of 1 and received a decrease in its trust value, it means that the user has shown signs of deterioration, reporting incorrect lineups. But since this user has such a high trust value, the participant still came to the conclusion to approving this users lineup based on the trust value of 0.9.

The next trust value that was tested was 0.2, by showing Figure B.9, as shown in Appendix B. The same reasoning was made by participant D as for the previous trust value. This time however, the participant would discard the users lineup due to that the value is so low and it spontaneously feels too low to approve, as the participant motivates the decision.

When going further, to test the next trust value, the most interesting part of this test session began. The next trust value that was to be tested, was 0.5. Figure B.6 in Appendix B shows the screen-shot displayed to the participant. The participant used the same reasoning as previously and came to the conclusion that this user's lineup most likely would be rejected. In conjunction to this, the participant had a wonderment, asking if this would be the process for each and every value existing, which was the case. The participant then said that, with the information available and by only being able to make a decision based on the trust value, the participant

would approve all lineups where the user has a trust value of 0.8 or above, since it feels like a safe and good value. Whilst incoming lineups, from users having a trust value of 0.4 or below would be rejected by this participant. From this, a great discussion arose regarding the trust value and its usage. The participant felt that it was hard to make a decision only based on the trust value, since a user who has a low trust value might have proven to be consistent lately, having the latest reported lineups being correct, and thus showing indications of improvement. For such a case, having a user with low trust value, the participant would reject the users lineup due to the low value. Whilst if there would be another value in conjunction with the trust value, providing an indication of whether or not the user is improving, the participant would most likely approve the users lineup, having a low trust value, but with an indication of consistency and improvement. The same goes for the other way around.

As for the second participant that took part in the this validation session based on users trust values, referred to as **Participant E**, was a bit more strict and reserved when approving or discarding lineups only based on the trust value. During the validation session, when showing the different screen-shots of the administrative user interface with users trust values, the only values that were shown until the participant made conclusions, was Figures B.10, B.9, B.6, B.1 and B.8, in the order they are presented. The participant mentioned that for all users having a trust value above 0.9, an automated decision of approving the lineup would be in place, since it most likely means that the user has reported many good lineups. But for all values below, the participant did not feel comfortable to make any concrete decision. For the low values, such as 0 and 0.2, the participant would discard the lineups, since that would either mean that the user only has reported very few good lineups, or that the user has reported many lineups, of which many of them have been incorrect. The participant would also not approve a lineup if a user has a trust value of 0.5, since it sounds very low, according to the participant. When continuing the discussions with the participant about the reasoning behind these decisions, this participant also mentioned that another value would be valuable next to the trust value, such as number of reported lineups, or that more values are involved in the algorithm behind the calculation of the trust value. The participant mentioned that, if a user's last five lineups have been correct, the participant would most likely approve that user's reported lineup. For such users, the participant felt that it would be good to, after five correct lineups in a row, give the users a higher increase in the trust value than only 0.1, since those users have proven to be consistent by providing correct and qualitative lineups.

The third participant that took part in this validation session, referred to as **Participant F**, had a completely different mindset than the previous two participants in this validation session. What is worth mentioning is that this participant has previous experience and knowledge about trust values and trust management systems. The participant felt in general that the trust value itself is not enough data to make any decision regarding a users lineup. The participant would, in conjunction to the trust value, like to in some manner compare the lineup for the same team,

from different users, in order to be able to validate if the reported lineup actually is correct. Because, if five users provide the same lineup, it is more likely that it is correct, than if only one user provides the lineup. Also, by only looking at the trust value for making an automated decision, the entire trust is put in the user only. And even though a user might have proven to be trustworthy based on previous performances, that does not necessarily have to mean that the next reported lineup by this user will be correct. This all comes down to the fact that this participant would not approve a users reported lineup based on the trust value only, no matter what the value is.

9.3 Iteration 2

For this iteration, another metric was added to the trust management system, which was the quality of each reported lineup. This means that the function providing each users' trust value, is now based on their corresponding history of reported lineups and the quality of each lineup. The quality of a lineup is viewed as the number of adjustments required to be made before the lineup can be approved in the administrative user interface. The number of adjustments acceptable so far in Forza Reporter, have been a maximum of three player positions or one incorrect player. This means that lineups sent to Forza Reporter, where a maximum of three players positions have been wrong or one player has been incorrect, have first been modified in the administrative user interface, and then approved and displayed in Forza Football. This means that there are five different qualities of lineups: correct, correct with one mistake, correct with two mistakes, correct with three mistakes and incorrect. For the four first cases, it means that the user's trust value now also should depend on the quality of the lineup. A completely correct lineup should still result in an increase of 0.1 to the trust value. As for the other lineups, the increase in the trust value should depend on the number of adjustments required prior to its approval.

As for the case of an incorrect player in the lineup, there have been changes made to Forza Reporter, since the brainstorming session was conducted with a co-worker at Forza Football AB. One of the features added since, was giving users the possibility of adding players that might be missing in a squad, or even adding an entire squad, in conjunction with reporting the lineup. As mentioned in Section 9.1, users had no option before when one player was missing in a squad, and were thus forced to put an incorrect player in order to be able to report any lineup for that particular team. In some cases, users would write directly to the organization, letting them know the mistake, or simply hope that the organization spot it themselves. Prior to this feature being added, this kind of mistake was in essence not a conscious one by the user, since they did not have any other option, which was the reason for why the discussion in the brainstorming session led to that these users should receive the same increase in their trust value as a user reporting a completely correct lineup. But since the addition of this feature, if a player is incorrect, the mistake is larger from the users side, meaning that the increase in the user's trust value should not be the same as for a user reporting a completely correct lineup.

The open-ended interview that was conducted, as presented in Section 8.4.1, provided insights in which order the different adjustments are ranked, as for their severity. The one considered to have the largest severity is the case of an incorrect player, meaning that this one should give the lowest increase to a users trust value. The reasons are given in the paragraph above. Besides this one, the adjustments acceptable are a maximum of three player positions. The more positions that are incorrect, the lower the increase to a users trust value should be. This additional metric, the quality of a lineup, was added to the function created in the first iteration. The function now takes, for each user, all reported lineups from the start until the last reported lineup by that user, and for each lineup checks whether it was correct or incorrect. For the former case, there are four possible outcomes: correct, which still gives an increase of 0.1 to the trust value, correct with one adjustment (which is one player being incorrect), giving 0.02 as an increase to the trust value, correct with two adjustments, giving an increase of 0.07 to the trust value, and correct with three adjustments, giving an increase of 0.04 to the trust value. For an incorrect lineup, the trust value is still decreased by 0.2. After scanning through all reported lineups for a user, the function outputs the sum, which provides the user's current trust value. This new metric that was added to the calculation of the trust value had to be evaluated. This was done in the same manner as in the first iteration, having participants make decisions about users lineups based on two criteria: users' trust values and users' histories of reported lineups. The results are presented in the following two sections.

9.3.1 Validation session based on users history of reported lineups

In Section 8.4.2, it was mentioned that the test session began in the same manner as previously had been described in Section 8.3.2. The difference in this test session was however that the participants got to see the users history of reported lineups together with a new value - the quality of each lineup. The results from this test session are presented in Table 9.2. Note that the results of the examples shown of a users history of reported lineups, corresponding to a certain trust value, are presented in an increased order of trust values, and not in the order in which they were shown to the participants in the test sessions. The first column refers to each history, corresponding to a certain trust value, that was used for the test session. The other two columns show the two possible decisions that the participant could make, which was to either approve or discard a users lineup based on the history. For each participant, the decision and the reason for the decision are presented in the table.

The history of a user as displayed in Appendix A, figure:	Approve	Discard
A.13 (Corresponding to a trust value of 0.06)		**Participant A:** Many discarded lineups and seems to have many adjustments on approved lineups. - - - - - - - - - - - - - **Participant B:** Too few data points and almost half of them have been discarded. - - - - - - - - - - - - - **Participant C:** No totally correct lineup.
A.14 (Corresponding to a trust value of 0.17)	**Participant A:** Approve, based on the last two being approved	**Participant B, C:** Only two approved lineups, with one and three adjustments respectively. Cannot be trusted.
A.15 (Corresponding to a trust value of 0.27)	**Participant A:** The user has more approved than discarded lineups. And two adjustments are not that bad. - - - - - - - - - - - - - **Participant B:** Similar reasoning as A.	**Participant C:** Too few data points to automatically approve.
A.16 (Corresponding to a trust value of 0.34)	**Participant A:** Good recent track record. Only one adjustment required for two lineups. - - - - - - - - - - - - - **Participant B:** Very much data with improvement by the user.	**Participant C:** The user shows an improvement, but would still not automatically approve.

Table 9.2: Results from second validation session, based on users histories.

Continuation of table 9.2		
A.17 (Corresponding to a trust value of 0.44)	**Participant B:** Users can make mistakes. The last five have been approved. User shows good improvement.	**Participant A:** User is not consistent. If you doubt, kick it out. - - - - - - - - - - - - - **Participant C:** Too many discarded lineups, where the first 7 are incorrect. Would not automatically approve.
A.18 (Corresponding to a trust value of 0.54)	**Participant A:** Approved based on good track record, even with adjustments.	**Participant B:** Hard to make a decision, only five lineups. The user used the app at two moments. Would rather discard than approve, due to few data points. Cannot trust the user enough. - - - - - - - - - - - - - **Participant C:** Only approved lineups but would not automatically approve the next lineup. If the user would only provide approved lineups in the future, giving more data points, would automatically approve the next lineup.
A.19 (Corresponding to a trust value of 0.66)	**Participant A:** Gives the user the benefit of the doubt due to good track of approved, and the last four are approved and did not need any adjustments. - - - - - - - - - - - - - **Participant B:** A user who did not get the concept in the beginning but approved.	**Participant C:** Starts with a lot of discards, then provides some approved lineups with adjustments and then some discards again. Does not quite seem to understand the concept or makes mistakes based on something else. Would not automatically approve.

Continuation of table 9.2		
A.20 (Corresponding to a trust value of 0.71)	**Participant A:** Things change prior to the start of a match, such as player positions. When clubs release the lineups they release it as a list at times, and some users still add the formation because it is cooler, based on previous lineups from the same team. - - - - - - - - - - - - - **Participant B:** A user who knows how to send in lineups even though there is one lineup with three adjustments.	**Participant C:** The user seems to have understood the purpose in the end, but not entirely. Shows indications of improvements, but not enough for getting a lineup automatically approved.
A.21 (Corresponding to a trust value of 0.86)	**Participant A:** Approve the lineup despite adjustment, based on good recent track record with 11/12 correct lineups. - - - - - - - - - - - - - **Participant B:** Very many data points. Only two lineups with adjustments.	**Participant C:** Would be good to know why the lineups were discarded, because it affects the decision. If it for example was discarded based on shirt numbers being incorrect, the decision would be more positive. But based on the situation, would not approve automatically.
A.22 (Corresponding to a trust value of 0.94)	**Participant A:** Approve based on that the user has more approved lineups and only one discarded. Three adjustments are probably only a mistake since the user has been consistent before. - - - - - - - - - - - - - **Participant B:** Similar reasoning as A.	**Participant C:** Would not approve it automatically, based on that the user has two incorrect lineups out of 12 lineups in total.

9.3.2 Validation session based on users trust values

Just as for the previous test sessions, this one was conducted with three different co-workers participating in one session each. As mentioned in Section 8.4.2, the test session started off, after giving each participant information about the test session, with showing a screen-shot of the administrative user interface with a user who has reported a lineup with a trust value of 0.71, as shown in Appendix B Figure B.19. The results from this test session are presented in Table 9.3. Note that the results are presented in an increased order of trust values, and not in the order in which they were shown to the participants in the test sessions. The first column refers to each trust value, that was used for the test session. The other two columns show the two possible decisions that the participant could make, which was to either approve or discard a users lineup based on the trust value. For each participant, the decision and the reason for the decision are presented in the table.

The trust value of a user as displayed in Appendix B, figure:	Approve	Discard
B.12 (Trust value: 0.06)		**Participant D:** Reject due to that the lineups reported by the user have most likely been wrong from the beginning. - - - - - - - - - - - - - **Participant E:** Value is too low.
B.13 (Trust value: 0.17)	**Participant D:** If only two lineups have been reported, it means progress, and it is therefore likely that the lineup is correct, but the participant would not approve it automatically.	**Participant E:** Most likely few reported lineups, way too few to approve the user's lineup based on the trust value.

Table 9.3: Results from second validation session, based on users trust values.

Continuation of table 9.3		
B.14 (Trust value: 0.27)		**Participant D:** The user probable has reported many lineups, but with many mistakes probably - unlikely to approve it. - - - - - - - - - - - - - **Participant E:** Too few lineups, and too low trust value to approve.
B.15 (Trust value: 0.34)	**Participant D:** Would not discard it.	**Participant D:** But would also not approve it. - - - - - - - - - - - - - **Participant E:** Would not approve it, value still too low.
B.16 (Trust value: 0.44)		**Participant D:** Very unlikely to approve it, but would not automatically rejecting it. - - - - - - - - - - - - - **Participant E:** Would not approve it, value too low.
B.17 (Trust value: 0.54)	**Participant D:** Likely to approve, but not automatically.	**Participant E:** Would not approve it, value too low.
B.18 (Trust value: 0.66)	**Participant D:** Likely to approve, but not automatically.	**Participant E:** Would not approve it, value too low.
B.19 (Trust value: 0.71)	**Participant D:** Would put the lineup as likely, but not automatically approve it.	**Participant E:** Would not approve it, value still too low.

Continuation of table 9.3		
B.20 (Trust value: 0.86)	**Participant D:** Spontaneous reaction is to automatically approve the lineup, due to that the trust value feels high and safe enough. - - - - - - - - - - - - - **Participant E:** Was a bit uncertain, but value feels high enough to approve the lineup.	
B.21 (Trust value: 0.94)	**Participant D, E:** Would automatically approve the users lineup.	

As for the third participant in this validation session, referred to as **Participant F**, the same reasoning was used as during the validation session conducted for the first iteration. The participant still felt that the trust value itself is not enough data to make any decision regarding a users lineup. The participant would, in conjunction to the trust value, like to in some manner compare the lineup for the same team, from different users, in order to be able to validate if the reported lineup actually is correct. By only looking at the trust value for making an automated decision, the entire trust is put in the user only. And even though a user might have proven to be trustworthy based on previous performances, that does not necessarily have to mean that the next reported lineup by this user is correct. This all comes down to the fact that this participant would not approve a users reported lineup based on the trust value only, no matter what the value is.

9.4 Iteration 3

As oppose to the two previous iterations, which were conducted in such an order that a metric first was added to the trust management system, implemented and then evaluated through validation sessions, this iteration was conducted by adding various metrics next to the trust value in order to evaluate how this may influence the participants' decisions. Since the results so far have indicated that the participants are not willing to completely rely on the trust value only, there was a need for evaluating the trust value along with other data available of all users, in order to evaluate if the trust value becomes significant enough for automated decisions first when another data point is added, or if the trust value is insignificant for the purpose of this study. Therefore, the values that were evaluated alongside the trust value

was: users approval rate, users' consistency, users number of reported lineups and users' approval rate together with the number of reported lineups. In order to be able to make conclusions regarding whether or not the trust value is insignificant for the purpose of this study, a validation session was also conducted using users' approval rate and number of reported lineups, only. Meaning that the trust value was excluded in this validation session. Note that all data that has been used in this iteration, is real data taken from existing users in Forza Reporter. The participants that took part in these validation sessions are the same that took part in the validation sessions for the two previous iterations, meaning that there were six participants in total for this iteration. In the following subsections, the results from each validation session is presented.

9.4.1 Evaluating the trust value together with users approval rate

The first metric that was evaluated alongside the trust value, was users' approval rate. The approval rate is a value that takes a user's all reported lineups, from the beginning until the last reported lineup, and indicates how many of all reported lineups have been approved, measured in percent. The results from this validation session is presented in Table 9.4.

The trust value of a user as displayed in Appendix C, figure:	Approve	Review lineup	Discard
C.1 Trust value: 0 Approval rate: 64.29%		**Participant B:** Would most likely review the lineup. If another lineups exists, this one would be rejected. - - - - - - - - - - - **Participant C:** Manual review, based on approval rate.	**Participant A:** Rather high approval rate, but gets worse by time, meaning that the last lineups have been rejected. - - - - - - - - - - - **Participant D, E:** Similar reasoning as A.

Table 9.4: Results from third validation session, based on users trust values and approval rate.

Continuation of table 9.4			
C.2 Trust value: 0.1 Approval rate: 66.67%		**Participant B:** Most likely two discards and one approved. For those with low trust value, if it is the only reported lineup, review it, otherwise discard it. - - - - - - - - - - - **Participant C, E:** Would not auto approve, feels like too few lineups. Would send it to review.	**Participant A:** Based on the low trust value. And the approval rate is not that convincing. - - - - - - - - - - - **Participant D:** Trust value is too low.
C.3 Trust value: 0.2 Approval rate: 75%		**Participant A:** Good value of approval rate, but very low trust value. - - - - - - - - - - - **Participant C, D:** Would neither auto approve or reject the lineup. Send it for manual review.	**Participant B:** Hard to tell but potentially too few data points, too low trust value, discard. - - - - - - - - - - - **Participant E:** The user has most likely only had 3 approved and one discarded. Feels like too few data points.

Continuation of table 9.4

C.4 Trust value: 0.3 Approval rate: 100%		**Participant A:** Send for review, a bit suspicious due to low trust value. - - - - - - - - - - - **Participant B:** Two approved lineups, would most likely review it. But if it is the only lineup, auto approve it. - - - - - - - - - - - **Participant C:** Send to manual review. - - - - - - - - - - - **Participant D:** Due to too few data points, send for review. - - - - - - - - - - - **Participant E:** Due to the high approval rate, would review the users lineups until the user at least reaches 0.5 in trust value, having the same approval rate.	

Continuation of table 9.4

C.5 Trust value: 0.4 Approval rate: 23.53%		**Participant B:** Bad approval rate but trust value is in around the middle. - - - - - - - - - - - **Participant C:** Send to manual review.	**Participant A:** Not good enough trust value and low approval rate. - - - - - - - - - - - **Participant D:** The user might be improving. Would like to see the users trend. Is the user improving or not in the last lineups? But most likely reject due to low approval rate. - - - - - - - - - - - **Participant E:** Feels like too few reported lineups, with a low approval rate. Would most likely reject immediately.
C.6 Trust value: 0.5 Approval rate: 80%		**Participant A, B:** Bases decision on the high approval rate. - - - - - - - - - - - **Participant C, D, E:** Would neither auto approve or reject the lineup. Send lineup for manual review.	

Continuation of table 9.4

C.7 Trust value: 0.6 Approval rate: 87.50%	**Participant A:** Bases on good enough values. - - - - - - - - - - - - **Participant B:** If 0.6 in trust value and 75% in approval rate, would sent to review. But due to the high approval rate in this case, approve. But depends a bit on if there are other lineups.	**Participant C, D, E:** Would neither auto approve or reject the lineup. Send it to manual review.	
C.8 Trust value: 0.7 Approval rate: 88.89%	**Participant A, B:** High enough trust value and approval rate.	**Participant C, D:** Would neither auto approve or reject the lineup. Send it to manual review. - - - - - - - - - - - - **Participant E:** Depends on what the last five lineups have been like. But based in this data only, would most likely send it to review.	

Continuation of table 9.4

C.9 Trust value: 0.8 Approval rate: 50%	**Participant A:** Only based on the trust value. Gives the user the benefit of the doubt.	**Participant B:** A user who has proven to be good lately, but half have been discarded. - - - - - - - - - - - - **Participant C, D:** Would neither auto approve or reject the lineup. Send it to manual review. - - - - - - - - - - - - **Participant E:** Review the lineup, in best case, since half have been incorrect.	
C.10 Trust value: 0.9 Approval rate: 100%	**Participant A, B, E:** Feels given to approve due to good track record with 8 correct in a row. - - - - - - - - - - - **Participant D:** Would definitely approve the lineup.	**Participant C:** Send for manual review.	

Continuation of table 9.4

C.11 Trust value: 1.0 Approval rate: 88.98%	**Participant A, B:** Based on really high trust value and a good approval rate. - - - - - - - - - - - **Participant D:** Improvements lately with a good track record. - - - - - - - - - - - **Participant E:** Would most likely approve it, but depends a bit on if the user has improved in the last lineups or if they have been rejected.	**Participant C:** Send for manual review.	

As for **Participant F**, who in the previous iterations could not make a decision regarding a user's lineup only based on the trust value, but rather wanted to compare the lineup for the same team between different users, still has the same mindset. The participant mentioned in this validation session that it is definitely easier to make a decision with the added metric, than only with the trust value. Despite this, the participant would still like to compare the lineups between different users as a precaution, since even though a user previously has proven to have a good track record, providing good and correct lineups, it does not mean that the next lineup from that user will be a correct one. It is not that important that the formations are the same, but at least that all players are the same.

9.4.2 Evaluating the trust value together with users approval rate and number of reported lineups

The second validation session was conducted by evaluating the trust value together with users' approval rate and number of reported lineups. The approval rate is, just like before, a value that takes a users all reported lineups, from the beginning until the last reported lineup, and indicates how many of all reported lineups have been approved. The number of reported lineups is value that represents a users total number of reported lineups, from the beginning to the end. The results from this validation is presented in table 9.5.

The trust value of a user as displayed in Appendix C, figure:	Approve	Review lineup	Discard
C.12 Trust value: 0 Approval rate: 64.29% #reported lineups: 14		**Participant A:** Send for review, based on decent approval rate. But the low trust value is a bit concerning. - - - - - - - - - - - **Participant C:** Send for manual review.	**Participant B:** Reject, but review this user's lineup sometime to give it a chance, since it might report a good lineup sometime. - - - - - - - - - - - **Participant D, E:** Not high enough approval rate and 0 in trust value = too bad values to be approved.

Table 9.5: Results from third validation session, based on users trust values, approval rate and number of reported lineups.

Continuation of table 9.5

C.13 Trust value: 0.1 Approval rate: 66.67% #reported lineups: 6		**Participant A, B:** A bit concerning about the low trust value and too few lineups. But the approval is above 50%. - - - - - - - - - - - **Participant C:** Send to manual review. - - - - - - - - - - - **Participant D:** Would review the lineup, but most likely auto reject it.	**Participant E:** Too few data points and low trust value.
C.14 Trust value: 0.2 Approval rate: 75% #reported lineups: 4		**Participant A, B:** Too few number of lineups reported and low trust value. But gives the user the benefit of the doubt based on approval rate. - - - - - - - - - - - **Participant C:** Send to manual review. - - - - - - - - - - - **Participant D, E:** Would neither auto approve or reject the lineups. Would send it for manual review.	

Continuation of table 9.5			
C.15 Trust value: 0.3 Approval rate: 100% #reported lineups: 2		**Participant A, B:** Too few number of lineups. Gives the user the benefit of the doubt based on not providing any incorrect lineups. - - - - - - - - - - - **Participant C:** Send to manual review. - - - - - - - - - - - **Participant D:** Too few data points. Will probably be correct, so would put it as likely but still review it. - - - - - - - - - - - **Participant E:** Review based on only two reported lineups (where both have been correct).	
C.16 Trust value: 0.4 Approval rate: 23.53% #reported lineups: 34		**Participant C:** Send to manual review.	**Participant A, B, D, E:** Many reported lineups, very low approval rate = reject immediately.

Continuation of table 9.5

C.17 Trust value: 0.5 Approval rate: 80% #reported lineups: 10	**Participant A:** Based on the approval rate of 80% and decent trust value.	**Participant B:** Would like to say approve, but due to the trust value, send for review. If it would have been 20 lineups with the same numbers, would say approve. - - - - - - - - - - - **Participant D:** Looks mainly at the trust value, seems like the user is getting worse. Would not auto approve. - - - - - - - - - - - **Participant C, E:** Send to manual review.	
C.18 Trust value: 0.6 Approval rate: 87.50% #reported lineups: 8	**Participant A:** High approval rate, good enough number of lineups. And the trust value is decent.	**Participant B, C, E:** Send to manual review. - - - - - - - - - - - **Participant D:** Send the lineup to review or auto approve. Feels rather positive.	

Continuation of table 9.5

C.19 Trust value: 0.7 Approval rate: 88.89% #reported lineups: 9	**Participant A:** Based on high approval rate, sufficient number of lineups. Good trust value. - - - - - - - - - - - **Participant D:** Seems good, would most likely auto approve.	**Participant B:** Good to review this one, due to 9 lineups only. Too few data points. - - - - - - - - - - - **Participant C:** Send to manual review. - - - - - - - - - - - **Participant E:** Depends on the trend (the last five reported lineups), but would send the lineup for review based on the info.	
C.20 Trust value: 0.8 Approval rate: 50% #reported lineups: 18		**Participant A:** Inconsistent user, would send it for review. - - - - - - - - - - - **Participant B:** Means that the user was worse in the beginning, but has improved lately. Review the lineup. - - - - - - - - - - - **Participant C:** Send to manual review.	**Participant D:** Under 70% in approval rate feels careless. Would most likely reject. - - - - - - - - - - - **Participant E:** Reject, since 50% of the lineups are incorrect.

Continuation of table 9.5			
C.21 Trust value: 0.9 Approval rate: 100% #reported lineups: 8	**Participant A, B, D, E:** Definitely approve, since all lineups have been approved and correct.	**Participant C:** Send to manual review.	
C.22 Trust value: 1.0 Approval rate: 88.98% #reported lineups: 127	**Participant A, B, D:** This is good, approve immediately. - - - - - - - - - - - **Participant E:** The user makes some errors, but would approve the lineup, based on a high trust value.	**Participant C:** Send to manual review.	

Participant F still has the same reasoning as before, meaning that the participant could not make any decisions about users' lineups based on the trust value, or with the help of the other metrics. The participant still wants to compare users' lineups with each other, in order to see if more users have reported the exact same players for the team, as a precaution. This is in particular of importance for higher leagues.

9.4.3 Evaluating the trust value together with the trend of the five last reported lineups

The third validation session was conducted by evaluating the trust value together with the trend of the last five reported lineups. This metric is a value that shows how many of the users five last reported lineups have been approved. Note that this metric does not tell the order of the five last lineups. Meaning that if a user has gotten 3 lineups approved and 2 discarded out of the five last reported ones, this metric will not give any indication of, if the discarded ones were the last ones or if the approved ones were so. The results from this validation is presented in Table 9.6.

The trust value of a user as displayed in Appendix C, figure:	Approve	Review lineup	Discard
C.23 Trust value: 0 #approved lineups out of five last reported: 3		**Participant A:** Review, too risky to approve. - - - - - - - - - - - **Participant C:** Send to manual review. - - - - - - - - - - - **Participant D:** Under 70%, so send it to review. If no one can review it, reject. - - - - - - - - - - - **Participant E:** Looks mostly at the metric of five last reported lineups.	**Participant B:** Interesting since the user has 0 in trust value, but 3/5 last lineups have been approved. Would say discard.
C.24 Trust value: 0.1 #approved lineups out of five last reported: 3	**Participant A:** Bases the decision more on the last reported lineups than the trust value.	**Participant B:** Review, since user shows improvements. - - - - - - - - - - - **Participant C:** Send to manual review. - - - - - - - - - - - **Participant E:** Review since the last 4/5 are correct lineups, despite low trust value.	**Participant D:** Too small trust value.

Table 9.6: Results from third validation session, based on users trust values and number of approved lineups out of the five last reported ones.

Continuation of table 9.6			
C.25 Trust value: 0.2 #approved lineups out of five last reported: 3	**Participant A:** Just based on having 3/5 correct lineups lately.	**Participant B, D:** Review, based on the "trend" (five last reported lineups). - - - - - - - - - - - **Participant C:** Send to manual review.	**Participant D:** Based on trust value, reject. - - - - - - - - - - - **Participant E:** Reject, due to that 2/5 have been incorrect and the trust value is low.
C.26 Trust value: 0.3 #approved lineups out of five last reported: 4		**Participant A:** Between review and just being approved. Torn between the last reported lineups. - - - - - - - - - - - **Participant B, C:** Send to manual review. - - - - - - - - - - - **Participant D:** More generous with low trust values when seeing the trend. Send lineup to review. - - - - - - - - - - - **Participant E:** Send to review. User might be improving.	

Continuation of table 9.6

C.27 Trust value: 0.4 #approved lineups out of five last reported: 3		**Participant A:** Not 100% sure due to the low trust value and 3/5 of the last have been correct. - - - - - - - - - - - **Participant B, C, D:** Send to manual review. - - - - - - - - - - - **Participant E:** Send for review, since the last 3/5 are correct lineups. Even though the trust value is low.	
C.28 Trust value: 0.5 #approved lineups out of five last reported: 5	**Participant A:** Based on 5/5 correct lineups lately. Users seems rather consistent.	**Participant B, C:** Send to manual review. - - - - - - - - - - - **Participant D, E:** Would not auto approve, feels like too few lineups. Send to review.	
C.29 Trust value: 0.6 #approved lineups out of five last reported: 5	**Participant A, E:** Approve since 5/5 are approved.	**Participant B, C, D:** Send to manual review.	

Continuation of table 9.6

C.30 Trust value: 0.7 #approved lineups out of five last reported: 5	**Participant A, B:** User seems consistent. - - - - - - - - - - - **Participant D:** Would most likely approve, due to good track record. Hard to decide without total number of reports though. - - - - - - - - - - - **Participant E:** Approve, due to improvements. Five last lineups are approved.	**Participant C:** Send to manual review.	
C.31 Trust value: 0.8 #approved lineups out of five last reported: 4	**Participant A, B:** Approve immediately. - - - - - - - - - - - **Participant D:** Feels that even if it is the last one that has been rejected but all of the others have been approved, it is a good track record.	**Participant C:** Send to manual review. - - - - - - - - - - - **Participant E:** Review, since 5/5 approved lineups would be auto approve for the next reported lineup, while 3/5 or 4/5 would be sent to review.	

Continuation of table 9.6

C.32 Trust value: 0.9 #approved lineups out of five last reported: 5	**Participant A:** Based on good values. - - - - - - - - - - - **Participant B:** If we can go in afterwards and decrease the trust value if the lineup would turn out to be incorrect. Hard to tell by only these two values. - - - - - - - - - - - **Participant D, E:** Approve immediately, due to good track record.	**Participant C:** Send to manual review.	
C.33 Trust value: 1.0 #approved lineups out of five last reported: 4	**Participant A:** Really high trust value and approval rate. - - - - - - - - - - - **Participant B, D:** Improvements lately, good track record. Would approve the lineup.	**Participant C:** Send to manual review. - - - - - - - - - - - **Participant E:** Would not auto approve if one lineup is incorrect out of the five last reported lineups. Send to review.	

Participant F has not changed the reasoning throughout the validation sessions, meaning that the participant still wants to compare users' lineups with each other, in order to see if more users have reported the exact same players for the team, as a precaution. The trust value is still valuable, but not enough as a basis for decision making regarding users' lineups, since a user who has proven to be trustworthy before, does not necessarily provide correct lineups in the future. So, according to this participant, it is required to make comparisons between the lineups and players as well.

9.4.4 Evaluating users approval rate together with the number of reported lineups

For the fourth, and final, validation session that was conducted in this third iteration, the trust value was removed. Instead, this validation session's purpose was to evaluate users' approval rates together with the number of reported lineups, in order to be able to make conclusions regarding the significance of the trust value. Both of the metrics have the same meaning as in the previous validation sessions conducted prior to this one. The results from this validation is presented in Table 9.7.

The metrics as displayed in Appendix C, figure:	Approve	Review lineup	Discard
C.34 #reported lineups: 14 Approval rate: 64.29%		**Participant A:** Approval rate is not that great. Decent number of lineups. - - - - - - - - - - - - **Participant B:** Apparently, a user who seems to like to throw in incorrect players at times. But send to manual review. - - - - - - - - - - - - **Participant C:** Send to manual review. - - - - - - - - - - - - **Participant D:** Under 70% approval rate. If no one can review it, reject it.	**Participant E:** Rather bad values.

Table 9.7: Results from third validation session, based on users approval rate and number of reported lineups.

Continuation of table 9.7

C.35 #reported lineups: 6 Approval rate: 66.67%		**Participant A:** The values are just above average, with an approval rate of almost 50/50. - - - - - - - - - - - **Participant B:** Would however discard it in case of other existing options. - - - - - - - - - - - **Participant C:** Send to manual review. - - - - - - - - - - - **Participant D, E:** Not that good values, but not horrible. Review the lineup.	
C.36 Approval rate: 75% #reported lineups: 4	**Participant A:** The values are decent.	**Participant B, C, E:** Send to manual review. - - - - - - - - - - - **Participant D:** One incorrect lineup. But is it in the beginning or end? Would send the lineup for review.	

Continuation of table 9.7

C.37 Approval rate: 100% #reported lineups: 2	**Participant A:** Suspicious, but would approve based on good start. The first lineups have been correct, so will take the risk. - - - - - - - - - - - **Participant B:** Too few data points, but user provides only correct lineups. Would approve it in order to make the system be fast. Rather give the benefit of the doubt, since we might find superusers.	**Participant C, E:** Send to manual review. - - - - - - - - - - - **Participant D:** It looks promising and is likely correct. But would still send for review.	
C.38 Approval rate: 23.53% #reported lineups: 34		**Participant C:** Send to manual review.	**Participant A:** A lot of lineups but approval rate is bad. - - - - - - - - - - - **Participant B, E:** Reject immediately - - - - - - - - - - - **Participant D:** Reject immediately. Would like to know if the last lineups are the correct or discarded ones. Easier to make a decision then.

Continuation of table 9.7

C.39 Approval rate: 80% #reported lineups: 10	**Participant A:** High approval rate and good number of lineups. - - - - - - - - - - - **Participant D:** Most likely correct, since 8/10 lineups are approved.	**Participant B, C, E:** Send to manual review.	
C.40 Approval rate: 87.50% #reported lineups: 8	**Participant A:** Auto approve. High approval rate and decent number of lineups.	**Participant B, C, E:** Send for manual review. - - - - - - - - - - - **Participant D:** Feels that it is only one incorrect out of 8 lineups, which is ok. Send for review, but most likely approve.	
C.41 Approval rate: 88.89% #reported lineups: 9	**Participant A:** Auto approve, good approval rate and decent number of lineups. - - - - - - - - - - - **Participant D:** Approval rate is above 70%. - - - - - - - - - - - **Participant E:** Approve, but not that obvious.	**Participant B:** Too few lineups for auto approve. - - - - - - - - - - - **Participant C:** Send to manual review.	

Continuation of table 9.7			
C.42 Approval rate: 50% #reported lineups: 18		**Participant A:** High enough reported lineups, but user is inconsistent. - - - - - - - - - - - **Participant B:** Closer to approve than discard, but send for review. - - - - - - - - - - - **Participant C:** Send to manual review. - - - - - - - - - - - **Participant D:** Feels like a messy person, review it.	**Participant E:** Reject, due to that 50% of the lineups have been incorrect.
C.43 Approval rate: 100% #reported lineups: 8	**Participant A:** User has good values. - - - - - - - - - - - **Participant B:** Few data points, approve until noticing that the user makes a mistake. - - - - - - - - - - - **Participant D:** Definitely approve, but would like to see the users trend. - - - - - - - - - - - **Participant E:** Approve, based on good track record.	**Participant C:** Send to manual review.	

Continuation of table 9.7			
C.44 Approval rate: 88.98% #reported lineups: 127	**Participant A:** Good number of lineups and decent approval rate. - - - - - - - - - - - **Participant B, D, E :** Would auto approve it. - - - - - - - - - - - **Participant C:** Too good values to not approve.		

Participant F, still feels that, even though it definitely is easier to make a decision with the added metrics rather than only using the trust value, the lineups between different users should be compared as a precaution, since even though a user previously has proven to have a good track record, providing good and correct lineups, it does not mean that the next lineup from that user will be a correct one. It is not that important that the formations are the same, but at least that all players are the same. Even though the participant values the comparison between lineups the highest, the participant would still like to use the trust value, but more as a threshold. For example, for a game in one of the highest leagues, a trust value of 1.0 is required. If comparing three users lineups, and the sum of their trust values is larger than the threshold, meaning 1.0, the lineup can be approved, choosing the user who reported the lineup first, providing them all an increase to their trust value.

9.5 Iteration 4

In the all of the iterations prior to this one, the trust management system has been in focus. Applying the knowledge base that was gathered in the beginning of this study, in order to design and evaluate the trust management system, in particular how to assign and use the trust value in crowdsourced information retrieval. In this iteration, the focus has been turned towards the process of feedback. This last iteration has had the purpose to evaluate what is good feedback to users participating in a crowdsourcing project and how significant the trust value is for the users, meaning whether or not the trust value should be known by the users or if other feedback is of greater importance to them. In order to evaluate this, the participants in the validation sessions in this iteration have been shown two different kinds of drafts of user profiles; on containing the trust value and one with a gamified element. The participants were first shown the former draft, and were asked the questions about the trust value as presented in Appendix D, Section D.1. They were thereafter

shown the latter draft, with the gamified element, and were asked the questions about the gamified element as presented in Appendix D, Section D.2. The interview conducted in conjunction to showing the participants the two drafts was open-ended, meaning that the participants were allowed to answer the questions as thoroughly and extensive as they wanted, opening up for further discussions. The results from these validations sessions, regarding what kind of feedback users should receive in a context similar to the one of Forza Reporter, is presented below.

9.5.1 Evaluation of whether or not feedback to users should include the trust value

The validation sessions conducted in this iteration started off by showing each participant Figure D.1 and conducting an open-ended interview, asking the questions about the trust value presented in Appendix D.2. In this section, the results retrieved from this open-ended interview are presented. The results from each question will be presented in their own table.

Would you as a user understand what that number means? If no, what further information would you like? Or is this not significant to you as a user?

Participant A:	It is not significant to a user. It looks like a low value without any further information. If the trust value would to be included, some sort of information box should be included, explaining what it means. The explanation could however be too complex for a normal user, and could thus discourage users.
Participant B:	I do not think the trust value should be visible to the user. It is something internal for the system to function. If a user sees that they have a low trust value it might have a negative effect on their dedication. However, showing something more arbitrary like reputation, level or achievements would be valuable for the user I think.

Table 9.8: Results from the fourth validation session.

Continuation of table 9.8	
Participant C:	The value is significant. When pressing on the element, the user should get more information about what the values means etc. It could however scare users if they get to see that they have low trust value. It could result in them stopping sending in lineups since they get the negative feedback. And we do not want to get stuck in a negative feedback loop. The feedback should be either positive or neutral. If the trust value is bad, do not say anything.
Participant D:	Would not understand what the value means, but would assume that it has something to do with the number of correct lineups provided by the user. Would rather like to see a value such as "Correct lineups" or "Correct lineup fraction", which is something a user immediately understands how to improve.
Participant E:	Would not understand the value by itself, as a user. It does not feel significant at all for the user. The feedback should rather be more something along the lines of letting the user know if it is trusted/not trusted. When pressing that element, more information should be provided about why the user is trusted/not trusted. It should furthermore also show how many lineups the user has reported and how many more are required to become trusted, if the user previously was not trusted.
Participant F:	Would guess what it means, but 0.4 does not mean anything to me. It sounds really low? I would like to know if I'm trusted or not. Perhaps I would prefer a rating (between 0-10, or between one and five starts).

Do you think that the user should receive some kind of feedback regarding its performances? If so, what kind of feedback do you think is of significance?

Participant A:	The feedback should be informative in such a way that it gives the user an indication of why a lineup has been rejected, and why a correct lineup has not been published.

Table 9.9: Results from the fourth validation session.

Continuation of table 9.9	
Participant B:	I think direct and quick feedback is great for showing the user that they have done something good. “Thank you for the lineup!”. Regarding performance feedback, I think it is important to make the user feel like they are contributing and focus on the positive aspects of the report, something like “Thanks to you, 53000 users knew that Ronaldo would start as a striker last weekend”. Improvement feedback might be important as well, maybe focus on helping the user understand rather than punishing them. “Woops, you reported an injured player. Do you know you can see injury status on X’s website?”.
Participant C:	For each lineup, feedback should be provided regarding whether or not the lineup was correct, and the the reason for it.
Participant D:	Yes, this is of great importance.
Participant E:	For each lineup, feedback should be provided. The feedback should include information about if the lineup was correct or incorrect and the reason for it. This especially for the cases when the lineup is correct, but not fast enough to be selected, and when the lineup is incorrect. This gives the user room for improving.
Participant F:	Yes, maybe something related to each reported lineup and also a general feedback like the rating. I think lots of people love making their "number" go up. Significant feedback would be whether the lineup was used, if it was correct and if not, what was wrong

For each rejected lineup, should the feedback include the new trust value, the reason for it being lowered, and an explanation for what the user has to do in order to reach a higher trust value again, or only the reason for rejection?

Participant A, Participant E:	Only the reason for rejection should be included in the feedback, along with what the user can do in order to improve. Does not think that the trust value is something that the user needs to see.

Table 9.10: Results from the fourth validation session.

Continuation of table 9.10	
Participant B:	I think the reason is enough. You might want to say something different depending on the severity of the error. Maybe they added 11 custom player names completely wrong, then we should try to be specific about this and explain what is wrong and how they can do better next time.
Participant C:	In this case, only the reason for rejection should be included. The trust value does not add any particular value to the user, since the user should not be aware of the negative consequences.
Participant D, F:	Only the reason for rejection is relevant for the user.

For each correct lineup, should the feedback include the new trust value and an explanation about what that new value means, or only the reason for the lineup being correct/published?

Participant A:	The trust value should not be included since it is not of significance for the user. In this case, it is only enough letting the user know if the lineup was correct/published, and why the a correct lineup potentially was not published.
Participant B:	I think the trust value should be hidden but make it abstract into achievements and levels. And reasons are always good if possible.
Participant C:	Include all information in the feedback to the user. Since the situation is positive, give the user as much feedback as possible.
Participant D:	Only the reason for the lineup being correct is relevant for the user.
Participant E:	Potentially if the user is not trusted, provide feedback about what needs to be done to become trusted. Otherwise, only the reason for rejection should be included in the feedback. The trust value is not something that the user needs to see.
Participant F:	Maybe the change in the value rather than the new value? or both. Also an explanation what the value means.

Table 9.11: Results from the fourth validation session.

Do you think that the trust value should be displayed to the user? Why/why not?

Participant A:	No, since the value could be interpreted as low when not having the entire background of it, potentially leading to discouragement of users.
Participant B:	I think not. The value does not say much to the user. It is very hard to relate as a user I think. Also, I think the trust value will be tweaked a lot during development and this might be super confusing for the user.
Participant C:	Yes, if the value is positive. If the value is negative, nothing should be shown to the user, besides the history of lineups and the reasons for them being rejected. But if the trust value is positive, the trust value is significant to show to the users, providing them feedback and incentives, by letting them know that they are on a good path.
Participant D:	No, not if it is not 100% clear what it means and how to affect it. The value could potentially be displayed to the user if there is some information about how it works exactly and that when you reach a certain value your lineups get automatically approved. It becomes a reason for fighting for something.
Participant E:	No, since the user will not understand the value. Further information could of course be provided, but the value still becomes insignificant to the user. The most important is if you are trusted or not, and what is required for a user to become trusted.
Participant F:	Yes, seeing the value imparts importance to the value. It ideally makes the user want to increase their trust.

Table 9.12: Results from the fourth validation session.

9.5.2 Evaluation of whether or not feedback to users should include gamification

Following the evaluation of whether or not feedback should include the trust value, was validation sessions for evaluating whether the feedback to users should include gamification along with the trust value, or if it completely should replace the trust value. These validation sessions started off by showing the participants Figure D.2 along with conducting an open-ended interview, asking the questions about the gamification element presented in Appendix D.2. In this section, the results retrieved

from this open-ended interview is presented. The results from each question will be presented in their own table.

Should the trust value be gamified? Meaning, if a user for example has a trust value of 0.1, the user is a water-boy. Or if the user has 1.0 in trust value, the user is a manager?

Participant A:	Definitely! This feels like something that could encourage users. It is however important that there is an explanation along with gamified value, letting the user know what it means.
Participant B:	Yes! And I think the level does have to connect to a specific trust value, it can also be dependant on the distribution of all users. So if top 10 trust values in the world are 'Grand master' instead of everyone with a trust value bigger than x.
Participant C:	Not a fan of the gamification. The product is serious, so no gamification should be integrated. Preferred would be achievements, meaning letting the user know how many line-ups have been reported, approved, rejected, the reasons for it etc.
Participant D:	Definitely! Especially if the user is able to collect badges/medallions for its achievements.
Participant E:	Better to include gamification than showing the trust value. If we need something more than only the trusted/not trusted element, (which was mentioned in by the participant in table 9.8), this would be good. The gamified value could also replace the trusted/not trusted element, as long as the user gets to know at what level it has become trusted.
Participant F:	I think it makes users want to report more and better. It would be one of the main drivers for users to report

Table 9.13: Results from the fourth validation session.

Should the trust value be displayed together with its gamification to the user? Or only the gamified value?

Participant A:	Trust value should never be shown. It is enough with the gamified value, only.
Participant B:	I think the trust value should be hidden for reasons mentioned in previous questions.
Participant C:	No gamification. Only trust value for those that are good.
Participant D:	Only the gamified value. No purpose in displaying both since they are more or less the same, just portrayed in different ways. The gamified value becomes more clear and fun for the user.
Participant E:	The trust value should never be shown.
Participant F:	Not sure. I guess the value will change more often (it is more granular) and feedback/change is important, so both.

Table 9.14: Results from the fourth validation session.

Should the gamified value be reflected from the trust value or from a different value? For example, the number of reported lineups? The number of approved lineups in a row? Or any other value that you can think of?

Participant A:	It can be based on the trust value, as long as the explanation of its meaning is there.
Participant B:	I see it like: Many factors affect the trust value (number of reports, percentage of approval, etc.) -> Trust Value and then we base the levels on this trust value.
Participant C:	Does not want to include any gamification.
Participant D:	Yes, it can be reflected directly from the trust value. Would be fun if the users could collect some kind of badges for each 5 correct lineups in a row, 10 correct lineups in a row and so on.

Table 9.15: Results from the fourth validation session.

Continuation of table 9.15	
Participant E:	The problem with if the gamified value is directly reflected from the trust value, it will be hard to get in enough levels. Would this be better to have the trusted and not trusted element and base the gamification levels on the number of approved/published lineups.
Participant F:	Definitely mostly from the trust value, but a combination might make sense.

Should the user receive feedback about what each gamified value means and what could that feedback be?

Participant A:	Yes, if you report a certain amount of correct lineups and reach a certain level, give the user feedback about that.
Participant B:	Yes, I think there could be a short description, hinting on what to think about in order to go to the next level, to incentive better lineups.
Participant C:	Does not want to include any gamification.
Participant D:	Yes! Would be nice with a “ladder”, showing the user all gamified values that exist and what the user needs to do in order to reach each value. Because then the users will know at all times what they are fighting for.
Participant E:	Yes, the user should receive that kind of feedback. When pressing on the gamified element, show a ladder of all possible gamified values that the user can reach, what the user needs to do to reach another value and why the user is on the current level.
Participant F:	No, it is fine if the value ("Waterboy" etc I am guessing) is a little vague.

Table 9.16: Results from the fourth validation session.

Should the user receive feedback about all the gamified values that exist and what the user needs to do in order to reach the next value in the hierarchy? If not, do you have another suggestion?

Participant A:	The user should receive information about all the values, in order to know what needs to be done in order to reach the next value.
Participant B:	I think you can say the next level maybe, but not all of them. I think the mystery of finding out what the next level is, is a great way to spark enthusiasm.
Participant C:	Does not want to include any gamification.
Participant D:	Yes! Would be nice with a “ladder”, showing the user all gamified values that exist and what the user needs to do in order to reach each value. Because then the users will know at all times what they are fighting for.
Participant E:	Show all levels or potentially only the next one and what is required to reach it.
Participant F:	It could go either way, but I think I prefer an unknown but guessable progression. Also, if your friends are "playing" you will see what their rank is and want to improve to reach the same rank as them. So there could be an indication about what the next rank is.

Table 9.17: Results from the fourth validation session.

Is there any feedback that you are missing? That you think would be of value to include?

Participant A:	Not in particular. The gamification element along with an explanation and feedback about the users’ performances is enough.
Participant B:	I think the bottom line is positive reinforcement, make people feel like they are a part of something bigger and be quick and direct when the users do good things, like reporting their first lineup etc. Rather be over-positive in the start of a user journey, than not giving enough.

Table 9.18: Results from the fourth validation session.

Continuation of table 9.18

Participant C:	It is hard to know what motivates the users, and thus a bit hard to tell what kind of feedback is good feedback. But what is crucial is to find more users that want to report at least a couple of lineups continuously, rather than one user that does everything.
Participant D:	Not anything in particular. The gamification element along with an explanation, a ladder of achievements ahead and feedback about the users' performances is enough.
Participant E:	Would be nice to make some sort of ranking for users for each team. For example, "*you are the number one reporter for FC Bayern München*". Another good feedback to users would be to add a ranking list of teams, so that users can go in and see which the top five reporters for each team are. If each gamification level requires for example 10 correct lineups, and if a user reports a lineup that gets rejected, the user should either go down to the level below or to the very bottom. Since a rejected lineup should be punished more than a correct one is awarded.
Participant F:	It would be good to include feedback about how quick I and how much quicker the quickest reporter was. It would also be good to provide feedback about how the user must improve to be the best. Another good feedback could be to let the user know how many people saw their reported lineup.

10
Discussion

The purpose of this study has been to explore possible ways to determine whether or not information retrieved through crowdsourcing from entities external to an organization is trustworthy. This explorations set the basis for further evaluations of how the trust can be assigned and then used for making automated decisions. In order to evaluate this, a trust management system has been developed during four iterations that were presented in the previous chapter. The system was developed for the Forza Reporter application, with the purpose of assigning trust values to users based on previous performances, in order to distinguish the trustworthy users from those who are not. The results from this study aim at answering the research questions which this study has revolved around. The organization Forza Football has a desire to in the future use these results in order to automate the decision of whether or not their users' reported lineups should be accepted.

This study has revolved around three main research questions, which previously have been presented in Section 1.1.

- **RQ1:** *Does using the trust value of a user to evaluate the trustworthiness of the information provided from that user yield similar results as a manual evaluation?*
- **RQ2:** *How should the trust value be used once assigned to the user?*
- **RQ3:** *How should the user receive feedback regarding what should be done to reach a certain trust value?*

In the coming sections, the results from the four iterations conducted in this study will be discussed in order to answer these three research questions.

10.1 The significance of the trust value

In this part of the chapter, the results from the two first iterations will be used as a basis for answering **RQ1:** *Does using the trust value of a user to evaluate the trustworthiness of the information provided from that user yield similar results as a manual evaluation?*. These iterations had the purpose of evaluating if using the trust value gives the same outcome as a manual evaluation would do. As has been mentioned in the literature study conducted prior to this stage, the trust value is an indication of users trustworthiness based on their previous performances. The trust

value can be both calculated and used in different ways. As presented in Section 3.2, there are four specializations of trust management systems; direct trust evaluation, reputation-based trust evaluation, socio-cognitive trust evaluation and organization trust evaluation [25]. The trust value can for example be given by other entities, where reputation is used in order to get an average trust value based on all of the other entities ratings, as described in 3.2.2. TripAdvisor and eBay are examples of organization adapting this kind of trust management system. This is something that could be of relevance for Forza Football in the future, when user profiles are added along with leader boards. Then, users could rate each others reported lineups, leading to that the overall rating of that particular user becomes its average trust value. As for the current situation, however, the direct trust evaluation model was the only one of relevance. Authors such as Wang and Singh [23], quantified the presence of the uncertainty in trust evaluation, and created an algorithm based on that, as presented in Section 3.2.1. The function calculates the uncertainty in a set of trust evidence, which are based on the distribution of positive and negative outcomes. The method provides a certainty value within a range [0, 1], which is based in statistical inference. The value 0 represents the highest value of uncertainty on the scale, while 1 indicates the opposite. In this study on the other hand, the trust value is based on users previous performances, increasing and decreasing the trust value with a set value, depending on if the lineup was correct or discarded, respectively. As mentioned, the trust value can be calculated in different manners, but it all depends on the context.

When evaluating the trust value in this study, three participants were asked to make a decision regarding users lineups only based on their trust value, while three completely different participants were asked to make a decision regarding users lineups based on their history of reported lineups. In the latter group, for the first iteration, both participant A and B would approve those users lineups whose history of reported lineups corresponded to a trust value of 0.4 or higher. In the second iteration, participant A would approve all users lineups, based on their histories, except for those whose history corresponded to a trust value of 0.06 and 0.44, while participant B would not approve the lineup for histories whose trust value corresponded to 0.06, 0.17 and 0.54. As for the third participant, C, who was more reluctant, who only gave one user a pass in the first iteration, approving the users lineup based on its history. This user's history of reported lineups corresponded to a trust value of 0.9, which was not known to the participant at that stage. For the other users, the participant either felt that the users rejection rates were too large or that the users had reported too few lineups in order to make a decision. In the second iteration, the participant would not approve any of the users' lineups, either due to too few data points, not knowing the reason for lineups being rejected or due to too few approved lineups out of the totally reported ones. This participant is rather on the safe side, than approving a users lineup despite the high trust value. The reason for the gap being so large between these participants could have something to do with their backgrounds and the various work that they do at Forza Football.

In the group making decisions based on the trust value, two of the participants, D

and E, did not have any prior knowledge or experience with the trust value, whilst participant F has previous experience from trust values, trust management systems and crowdsourced information retrieval. This participant was completely reluctant, in both iterations, against making decisions regarding users lineups only based on their trust value, since the trust value alone is not enough data in order to make a decision. This due to that, according to the participant, the trust value gives an indication regarding how trustworthy the user has proven to be previously, but it does not necessarily mean that the user will be as trustworthy when providing lineups in the future. This participant would thus not completely rely on the trust value for making a decision. As for the other two participants, their reasoning was more or less similar to each other. Both these participants in general felt that if they would have another value alongside the trust value, the decision making would be easier. Participant D would for example like to have another value that indicates if the user is improving or not, in its latest lineups. This way, if a user has a low trust value, but shows improvements in the recent lineups, the participant would most likely approve the users lineup. On the other hand, if only the low trust value is available, the participant would not approve the user's lineup, taking the risk of potentially displaying an incorrect lineup in the application. As for participant E, the most valuable metric that this participant would like as an addition alongside the trust value, would either be an indication of the users performances in the last five reported lineups or the total number of reported lineups.

When comparing these two groups, it is apparent that the participants making decisions based on users histories, were more determined and more secure in making their decisions. Whilst the group making decisions based on the trust value struggled with actually even making a decision, and explicitly expressed that they at least wanted another value alongside the trust value, which in some manner could indicate whether or not the user is improving in its latest lineups. One of these participants even wanted to take it a step further, by also comparing users' lineups with each other, as a precaution. The participants making decisions based on users histories, would approve lineups from users whose histories correspond to a very low trust value, whilst the other group would not do so. This therefore indicates that, even though the trust value may be of significance, the trust value alone is not enough in order to decide if an incoming lineup from a user can be approved and displayed in Forza Football. Even though it previously has been proven that the trust value alone can be used as a replacement for manual evaluation, for evaluating the trustworthiness of the information provided by users, it may not be applicable for all contexts. In the context of Forza Football, where the quality of the information is highly values, the trust value alone does not become trustworthy enough from the organizations perspective. To rely completely on one value, which indicates users performances in the past, but no indication of what the quality of the lineups in the future will be. Thus, the answer to RQ1 is: **No, the usage of the trust value of a user to evaluate the trustworthiness of the information provided from that user does not yield similar results as a manual evaluation.**

In the context of Forza Football, the fact that the trust value alone has been proven

to not yield similar results as a manual review, indicates that either more values are required alongside the trust value or that the trust value completely needs to be replaced with something else, in order to yield similar results as a manual review. As it was mentioned in the beginning of this section, the trust value can be calculated in different manners, but it all depends on the context. This also means that the usage of the trust value becomes context-dependent. In Section 3.1, it was mentioned that trust may have a different meaning depending on the persons involved and the context. Therefore, the fact that the trust value proved not to be of the expected significance in the context of Forza Football, does not necessarily mean that it will not be of significance for others. By not yielding similar results from the evaluations of the trust value as the manual reviews, could either depend on the persons involved in the validation sessions, or the context, which is the most important aspect in this case. Therefore, the conclusions that can be drawn, is that the approach used for this study is adaptive in contexts similar to Forza Football. It could most certainly be adaptive for other context as well, but for which, it is not safe to say.

The results from these two first iterations, led to that the necessity of a third iteration arose. In this third iteration, the trust value was evaluated alongside other metrics, in order to evaluate how to use the trust value and what decisions can be made of it, which will be the topic of discussion in the next section.

10.2 The usage of the trust value

In the literature, algorithms were found for different kinds of trust management systems, such as the Eigentrust algorithm [27], the Beta Reputation System [28], Wang and Singhs evaluation model [23], the socio-cognitive evaluation model proposed by Falcone and Castelfranchi [30] and the organizational trust evaluation approach proposed by Kollingbaum and Norman [31], but there is not that much literature about how organizations adopt trust management systems, and in particular, how they use the trust management system. TripAdvisor and eBay are two well-known organizations, which adopt the reputation-based evaluation model, providing some information about how their rating systems are designed. Even though both of the organization provide information about how their rating systems are designed and used, there is still a lack in the literature about how the internal processes work. Meaning that the actual use of the above mentioned algorithms, or trust management systems in general, is not something discussed widely. Further organizations that have been studied, known for adapting crowdsourced information retrieval, and some even known for using trust management systems, do not provide any information about how they manage it. Organizations that were contacted for interviews about their crowdsoured information retrieval projects, and in particular, how they adopt their crowdsourcing and trust management systems, did not show any desire to participate. This could be due to that organizations do not want to give out their secrets, to let other organizations know how they have solved their issues with crowdsourced information retrieval, and how some of them have integrated trust management systems and the use of trust values. This topic of discussion for this

section will this be to answer **RQ2:** *How should the trust value be used once assigned to the user?*.

In order to get a better understanding of how the participants in this study would like to use the trust value, the trust value was evaluated alongside additional metrics, which were presented in Section 8.5. Throughout all validation sessions in this iteration, participant C and F were still rather reluctant towards the trust value and could not make any decisions regarding users' lineups based on the trust value only. Participant F felt that, even though it definitely was easier to make a decision with the added metrics rather than only using the trust value, the lineups between different users should still be compared as a precaution, since even though a user previously has proven to have a good track record, providing good and correct lineups, it does not necessarily mean that the next lineup provided by that user will be a correct one. In the comparison between users' lineups, it is not that important that the formations are the same, but at least that all players are the same. Even though participant F values the comparison between lineups the highest, the participant would still like to use the trust value, but more as a threshold. For example, for a game in one of the highest leagues, a trust value of 1.0 is required. If comparing three users lineups, and the sum of their trust values is larger than the threshold, meaning 1.0, the lineup can be approved, choosing the user who reported the lineup first, but providing them all an increase to their trust value.

This participant has been reluctant towards the trust value throughout all iterations, and has had reasonings that deviate from the rest of the participants. One of the potential reasons could be that this participant is the only one who has previous experiences with trust management systems. Since this participant has worked with trust management systems in a particular way, where the automated decision making has not been based on the trust value directly, it becomes a difficult task at hand for the participant to reason in a different way. The participant has for several years worked with a systems that sets a threshold for which users' sum of trust values needs to exceed, and comparisons between users' data before making a decision, is the basis of the system. When working with a higher level of security, where users data is compared for making an automated decision rather than the trust value directly, the likelihood of the participant changing its mindset, to use the trust value directly, becomes relatively low.

Participant C on the other hand would never base a decision regarding a lineup on metrics that provide indications of users previous performances, especially not for higher leagues. One exception was however made in the fourth validation session in this iteration, where the participant would automatically approve that users lineup whose approval rate was 88.98% and the number of reported lineups was 127. This user had a trust value of 1.0, which was not known to the participant during this validation session. The participant further mentioned that, in order to even consider the function of automatically approved lineups, the participant would like the user to have at least 15 reported lineups, 1.0 in trust value and 100% in approval rate, meaning that the participant could consider automatically approved lineups above

a certain threshold. If a user would have 0 in trust value and an approval rate under 10%, the participant would automatically reject the lineup. The participant would however not even like to have the auto reject function at all, since a user who might be improving, gets rejected all the time due to a low trust value, or based on other metrics, and thus never gets the chance to improve. *For all other lineups, the participant would send them to manual review, since the participant only trusts a human in making a review of the current lineup and basing a decision based on its quality, rather than using a value indicating users previous performances.* This is rather important, since it provides an indication that this participant would require manual reviews for the majority of the incoming lineups, and is thus reluctant towards an automated process based on the trust value, or even other potential metrics.

Continuing, the participant felt that the best values for making a decision regarding users lineups, were the trust value and number of approved lineups out of the five, last reported ones. Despite this, the participant furthermore explicitly expressed that the risks are too large in basing a decision about a lineup on such a value, without reviewing the actual lineup. And the risk of potentially displaying an incorrect lineup, is not worth taking. As was presented in Chapter 3, one of the most important aspects of trust is that it involves risks. Taking risks means diving into the unknown. When interacting with other entities, these risks arise, since we need to completely rely on an uncontrollable situation, without any certainty of the final outcome [9], [18], [19]. For the case of Forza Football, this will at times be required, to dive into the unknown despite the risks, and despite the fact that some may be reluctant towards it. Because one of the current issues in Forza Football is the lack of data, especially for the lower leagues, which is the reason for the need of Forza Reporter and the crowds' knowledge. And for these lower leagues, there is sometimes no way of confirming if the lineup actually is the official one, in comparison to the higher leagues where there are official sources that the lineups can be compared to. For such scenarios, the trust value, or at least some value indicating users performances in the past, becomes crucial.

As for the remaining participants, when comparing the different validation sessions to each other, and comparing the participants' decisions based on the different metrics, it is evident that the participants have been rather consistent in some cases, providing the same decisions, based on different metrics. Whilst there are some evident cases where the participants' decisions have completely differed, depending on what kind of information has been available to them. One interesting example is trust value 0.1, looking at the decisions made by participant A throughout the four validation sessions. The participant would automatically discard that users lineup based on the trust value and approval rate, send the lineup for review based on the trust value, approval rate and number of reported lineups, automatically approve the lineup based on the trust value and the trend of the five last reported lineups, and send the lineup for review based on the approval rate and number of reported lineups only, which conforms to the decision made by the participant in the second validation session. This provides an indication that, the more data that is available for making a decisions, the easier the decision is to make. When the participant was

asked the two questions mentioned in 8.5, which were: "*Which of the metrics made the decision the easiest?*" and "*Do you think that this metric is enough for making a decision regarding a lineup, or would you still like to have the trust value?*", the participant felt that the most valuable metrics for making a decision regarding a lineup was: the trust value, number of reported lineups and approval rate, together. The participant felt that the trust value puts another dimension into it and gives more factors for making a decision, and would thus still include the trust value. The participant would also like to, as a precaution, compare the lineups between users, especially for higher leagues. This due to that, if several users have reported the same starting eleven for the same team, it is more likely that the lineup is official and correct, rather than if only one user has reported the lineup.

Participant B was more consistent in its decisions, in comparison to participant A, managing to in some cases make the same decision, no matter which additional metric was used. There are some scenarios where the participant made completely different decisions based on the different metrics, as presented in the tables in Chapter 9.4. This participant felt that the metrics that were most valuable for making a decision regarding users lineups was the trust value together with the approval rate. The participant further expressed that the trust value is of significance since it is a value that changes with the same values all the time, providing indications of how the user has performed in the past, and lately.

Participant D was also rather consistent, but had some values for which different decisions were made. In some cases the participant wanted to send to lineup to review due to uncertainty, which could depend on that the metrics available were not good enough information for making a decision, or that the participant was lacking some kind of metric alongside the others, for facilitating the decision making. For the decisions made from this participant depending on the different metrics, it is referred to the tables in Chapter 9.4. As for which of these metrics was considered to be the most valuable one in the decision making, participant D felt that the approval rate was the most interesting one. This participant expressed however, that it would be valuable to view the trend of the user, and suggested having a visual "trust value" represented in a matrix. This matrix would show the user's all reported lineups as green and red dots, where the green dots represent correct reported lineups and the red one represents discarded lineups. This would give the possibility to see the users overall trend, if the user improving or not.

Participant E was quite similar to participant D, being rather consistent throughout the various validation sessions, making similar decisions despite various metrics. There were some scenarios however where the participant did make different decisions, just like the other participants. All of these results are presented in Chapter 9.4. The value of largest significance or this participant was the metric of number of approved lineups out of the five last reported ones. This participant would like to see the trend of the user, either the last 5, 7 or 10 reported lineups. This means that the participant would like to see how many of the last 5, 7 or 10 lineups have been approved. The participant is not quite sure of how many would be most suitable,

but would like to see at least the five last reported ones. The participant would like to base the trust value on the last number of reported lineups instead, since it gives an indication of the user's improvements. If basing it on the five last reported lineups, the user is required to get 5/5 lineups approved in order for the process to be automated. For 4/5 or 3/5 correct lineups, the user's lineup should be sent for manual review. And anything below this should be discarded. Since it is required to have 5/5 correct lineups in a row in order to get the lineup automatically approved, it was discussed that if one incorrect lineup should lead to the user having 0 in trust value, having to start completely over. And if the user once again reports five in a row, the user has reached 1.0 in trust value, meaning that the lineup can once again be automatically approved. This means that in this particular scenario, when 5/5 correct lineups in a row are required, the trust value would increase with 0.2 for each correct lineup.

Even though the participants' reasoning has differed, it is evident that at least the majority of them find the trust value to be usable in some manner. Some of them have expressed that the latest lineups are of greatest significance, and that the trust value thus should be restricted for a certain time period or to a certain number of lineups. One of the participants expressed the need for a matrix, providing a visualization of the trust value, giving an immediate indication at a single glance regarding the users trend, meaning how the user has performed overall. Some of the participants also expressed the need for comparing users' lineups to each other, since the likelihood that a lineup is correct becomes larger when several users have reported the exact same lineup, or at least the same players in the starting eleven. One of these participants further gave the suggestion of using the trust value as a threshold when comparing users lineup. If the users' trust values exceed the threshold, and if the comparison between their lineups provides the same results, the first correctly reported lineup can be approved and displayed in Forza Football. All of the users should of course get an increase in their trust value, since all of them have reported the correct lineup. But the user who reported the lineup first, should receive a slightly larger increase in its trust value in comparison to the rest. The participants might have given various suggestions to how the trust value can be used, but have provided similar patterns, from which conclusions can be drawn. For larger and more popular leagues, the quality of the data becomes of even greater importance, meaning that for these leagues, the comparison between users' lineups becomes crucial, using a threshold for which the users' trust value needs to exceed in order for the lineup to be automatically approved. Since many of the participants expressed that they value the latest lineups more, rather than those that were reported several months ago, the trust value should be based on the latest lineups, and not on all lineups reported throughout the application's existence. What the exact number of lineups should be, may it be the last 5, 7, 10 or 20 lineups, is something that needs to be evaluated in the environment of Forza Reporter.

Thus, the answer to **RQ2, How should the trust value be used once assigned to the user?**, is: for the cases where quality is of greatest significance and incorrect data displayed to the entire crowd is of highest severity, the comparison between

users' data becomes crucial. In those cases, a threshold can be set for which users' trust values needs to exceed in order to get their data approved. This means, that for those cases where the data is extra sensitive and when it is crucial to be 100% sure that the data being displayed is completely accurate, larger cautiousness is required, meaning that relying on the trust value only, for making automated decisions, becomes insufficient. For lower quality controls, where the data coverage might not be as high, the trust value alone is significant enough for being used as a basis for automated decisions making.

What this in essence means for Forza Football is that, for higher leagues where the quality is of greatest significance, users reported lineups need to be compared to each other for the sake of cautiousness. For those cases, a threshold will be set for each league, for which users' trust values need to exceed in order to get their lineup approved, given that the comparison between their lineups provides the same results. For lower leagues however, where the data coverage is far from as large as for the higher leagues and where there is no way for Forza Football to determine whether or not a given lineup actually is official, as oppose to the larger leagues where official sources are used, the trust value can be used directly. Meaning that the trust value of a user can be used for making automated decisions in lower leagues. The trust value will be calculated in the same manner as in this study, but will be restricted for a certain number of lineups or a given time period.

10.3 Feedback to users

As was presented in Section 1.1, one of the aims of this study has been to provide practitioners with proper ways of providing feedback to their peers, in particular to those in the same position as Forza Football. This since larger parts of existing research provide guidance for how to give feedback to students, co-workers etc., but not in particular how to provide feedback to users taking part in a crowdsourcing application, that uses a trust management system, meaning should the trust value be displayed to users or is the value insignificant to them? Or should the value be gamified in some manner, adapting it more to the users and making it more understandable to them? Or should the feedback simple not include this at all? These are examples of questions which existing literature does not answer. Therefore, the section aims at using the results from the fourth iteration as a basis for answering **RQ3: How should the user receive feedback regarding what should be done to reach a certain trust value?**.

In the validation sessions regarding the feedback to users, the same six participants were included, as in the previous iterations. Since the feedback is directed towards the users of the application, it would have been appropriate to include some of the users in the validation sessions as well. Since it is the users who themselves know the best what motivates them to participate in reporting lineups, it is also they who know the best what kind of feedback they would appreciate. The most suitable users for this purpose, would be users of Forza Reporter, who continuously have reported lineups to the organization Forza Football. Since these users might be spread out

geographically, it would be a hard task at hand to meet with these users in order to conduct these validation sessions. It could however have been possible to conduct it via the web, even though it would have been preferred to do it in a physical meeting. Furthermore, the time constraints of this study, provided limitations for which participants could be included, made it impossible to include the users in the validations sessions. However, during this study there has been a parallel ongoing research internally at Forza Football, where users of Forza Reporter have been contacted in order to receive a better understanding of what motivates them to report continuous lineups, which has provided some insights into the motivations of users. Also, the validation sessions conducted with co-workers at Forza Football, has provided further insights into what kind of feedback should be provided to users, which also has proven to be aligned with the results retrieved from the parallel research conducted internally at Forza Football. Furthermore, the way that the validation sessions were constructed in this iteration, was in order to evaluate whether or not the trust value should be included in the feedback to users and what that feedback might consist of. And since the trust value is used for internal processes, whose meaning the users do not have any knowledge about, it is a decision that Forza Football needs to make, of whether or not to display the trust value to users, which also is the reason for conducting these validation sessions with co-workers at Forza Football. If however, users would be included in the validation sessions in this iteration, the validation sessions would have to be completed differently.

During the validation sessions conducted in the fourth iterations, the participants were shown some drafts and an open-ended interview was also conducted. One of the most important aspects of the evaluation in this iteration, was to evaluate of whether or not the trust value is of significance to include in the feedback to users. For Forza Football, the results imply that the trust value should not be displayed to the user, since it is used for internal processes and thus becomes insignificant to them. Instead, the user should receive feedback that is direct and quick, letting the user know the reasons for a lineup being correct but not displayed in Forza Football and the reason for rejection, giving the user room for improvement. The user should certainly also receive feedback when the lineup is correct and fast enough to be displayed in Forza Football. In order to motivate users to provide more and better lineups, from the users perspective, it would be more interesting to receive some kind of reputation through achievements, levels or similar. Even though the trust value itself is not significant to show to the user, which most likely is the reason for why other organizations in the same situation do not display it, other values reflected from the trust value could be, such as a gamified value. The gamified value becomes something representative for the trust value, but more adaptive to the users' language and thus more understandable. In the case of using a gamified value, the user should receive feedback about what meaning the current value has, what the next value is, and what the user needs to do in order to reach it, in order to motivate the users.

The general findings retrieved from the results in this study, gives the following answer to RQ3: How should the user receive feedback regarding what should be done

to reach a certain trust value?: **The user should receive feedback based on values familiar to the user, derived from the trust value, and not the trust value directly**. The trust value is a value used for internal processes, and since users neither understand it nor have any value of it, it is insignificant displaying it. The most important feedback is the one with positive reinforcement, giving users immediate and quick feedback. The feedback should include information about why the data was approved or rejected. If also using a value derived from the trust value, the feedback should include information about the current value, the next value and what the user needs to do in order to reach it.

10.4 Threats to Validity

When conducting a qualitative research study, there is a need to evaluate to what extent the research findings are believable and accurate, and what factors might affect its validity [64]. There are numerous approaches available for this particular purpose, but in order to find potential threats to validity in this study, Creswells's definition of internal, external and construct threats to validity has been used [64]. Internal threats to validity are procedures used in the study that might cause a threat for the researcher to draw correct conclusions from the data retrieved from the evaluations. External threat to validity is considered to arise when incorrect generalized conclusions are drawn from the data retrieved from the evaluations in the study.Construct validity concerns potential incorrect measurements used in a study.

10.4.1 Threats to Internal Validity

One potential internal threat in this study could have arouse due to this study's, participants' previous experiences and current work positions at Forza Football. A total of six participants were a part of the evaluations taking place in this study, among which two of them work as data specialist, three frontend developers and one backend developer, with previous experiences in trust management systems and trust values. Since the latter participant has worked with trust management systems previously, particularly in a certain way, where the automated decision making has not been based on the trust value directly, it becomes a difficult task at hand for the participant to reason in a different way. The participant has for several years worked with a systems that sets a threshold for which users' sum of trust values needs to exceed, and comparisons between users' data before making a decision, is the basis of the system. When working with a higher level of security, where users data is compared for making an automated decision rather than the trust value directly, the likelihood of the participant changing its mindset, to use the trust value directly, becomes relatively low. This participant was reluctant towards the trust value throughout all of the evaluations, providing rather few data points for the evaluations.

Furthermore, one of the participants working as a data specialist considered the quality and correctness of data to be of highest priority, and would thus never rely on a certain value to be used for making automated decisions regarding users line-ups. This participant was in other words also reluctant towards the trust value.

Using participants with different backgrounds, can be viewed from a two-sided perspective. On one hand, it can provide a wide data set, and thus covering more aspects, providing different viewpoints to the problem at hand. On the other hand, this can limit the ease of drawing generalized conclusions, due to the data collected being so wide, consisting of completely different opinions. However, each individual reasons in its own way. So to use participants with the same background and experiences does not necessarily mean that the results in this study would become more accurate and thus enable one to draw generalized conclusions.

10.4.2 Threats to External Validity

One potential external threat to validity could have arose in this study due the number of participants used in the evaluations. The number of participants that have taken part in the evaluations in this study has been a total of six participants. Also, these participants have all been co-workers at Forza Football, and have thus provided data points for the given context. This has enabled to draw conclusions for the problem at hand in the context of Forza Football, but it is not safe to draw too generalized conclusions from the data retrieved in the study. For the context of Forza Football, the number of participants did not affect the results negatively in any manner, since the participants were carefully chosen, selecting those participants who normally make decisions regarding both Forza Reporter and Forza Football. A further explanation is made in Section 8.3.1. But for drawing generalized conclusions from such a small data set, might not be accurate.

10.4.3 Threats to Construct Validity

An aspect in this study that could be viewed as a construct threat to validity, is how the trust value has been calculated. The trust value can be calculated in a mathematical manner, using models and formulas, or the way it was done in this study. The more accurate way, providing a better foundation for arguments, is using mathematical models and calculations. However, for the purpose of this study, this has not quite been necessary. One of the aims of this study has been to evaluate if and how trust can be used in order to distinguish trustworthy users from those who are not, in order to make automated decisions. And thus, the importance has not in particular lied in how the trust value is calculated. For the purpose of this study, and in order to conduct the study in a proper way, answering the research questions that are the basis of this study, the way the trust value is calculated has not been an affecting factor for the outcome of this study.

11
Conclusion

The purpose of this study has been to explore possible ways to determine whether or not information retrieved through crowdsourcing from entities external to an organization is trustworthy. To enable this, a trust management system was developed with the purpose of assigning trust values to users based on previous performances, in order to distinguish the trustworthy users from those who are not. The study was a collaboration with the organization Forza Football AB who has a desire to use the results from this study in order to automate the process of approval and rejection of users' lineups.

The scientific contribution of this study has been to provide other practitioners, such as application developers adapting crowdsourcing, guidance in what various ways users can be distinguished using trust, in particular by integrating trust management systems into their current systems. The study has had the aim of providing these practitioners with suggestions about how trust management systems can be integrated into crowdsourcing applications and how the trust value can be used in various ways in order to distinguish trustworthy users from those who are not.

The study has further had the aim of providing these practitioners with guidance in what proper feedback in this context may be. The study has aimed to evaluate if the trust value should be displayed to users or if it is insignificant to them. If the value should be gamified in some manner, being more adapted to the users and making it more understandable to them, or if it should not be included at all in the feedback given to users.

This study has in particular had the aim at answering the following research questions:

- **RQ1:** *Does using the trust value of a user to evaluate the trustworthiness of the information provided from that user yield similar results as a manual evaluation?*
- **RQ2:** *How should the trust value be used once assigned to the user?*
- **RQ3:** *How should the user receive feedback regarding what should be done to reach a certain trust value?*

One of the main findings in this study was that the usage of the trust value of a user, to evaluate the trustworthiness of the information provided by that user, does not yield similar results as a manual evaluation (**RQ1**). This, along with the liter-

ature review conducted in this study, further provided the finding of that the trust value is context-dependent. Thus, the fact that the trust value proved not to be of the expected significance in the context of Forza Football, does not necessarily mean that it will not be of significance for others. The conclusion that safely can be drawn for the first research question, is that the approach used for this study is adaptive in contexts similar to Forza Football.

Another finding in this study was how the trust value should be used once assigned to the user (**RQ2**), in the given context. The evaluations conducted for answering this research question, provided mixed opinions from the participants. Some wanted to use the trust value directly, but in the case of it being restricted to the last reported lineups or to a certain time period. Another participant wanted the trust value to be visualized, showing all reported lineups by the user in a matrix, where the correct lineups are represented by green dots and rejected lineups are represented by red ones. By doing so, the participant's performances could be viewed in a single glance. The remaining participants did not want to rely on the trust value for making an automated decision, but rather wanted to compare the lineups between users in order to see if they have reported the same. These participants wanted to set a threshold for all leagues, for which users trust values would be compared to. If the comparison between the users' lineups proved to yield same results, and if their trust values exceeds the threshold, the lineup can be approved. This evidently shows that the trust value has many applications and that it can be used in multiple ways. These conclusions are in the context Forza Football, but the generalized conclusion that can be drawn is that, when high quality controls are required, the comparison between users data and setting a threshold for the lowest trust value required for getting data approved, becomes applicable. Since the information displayed to other users can be very sensitive, the importance of the data being accurate, increases the importance of being cautious, meaning that automated decisions based on the trust value only, is not trustworthy enough. For lower quality controls, where the data coverage is lower, or where there is no way to compare users data, the trust value can be used directly.

The final finding in this study was that the feedback to a user always should have a positive reinforcement, providing users immediate feedback, letting them know why their data has been approved or rejected. For the latter case, it is very important to let the user know the reasons for the rejection so that they can improve until next time. Another important finding in the evaluation of feedback to users was that the trust value should not be included in the feedback to users at all (**RQ3**). This is a value used for internal processes, and since the users neither understand it nor have any value of it, there is no need in displaying it. Instead, values more familiar to users, derived from the trust value, can be included in the feedback, such as gamification. When using such a value, in order to motivate users, it is of great importance to provide feedback both about what their current value means, but also to provide feedback about what the next value is and what the user needs to do in order to reach it.

References

[1] B. Dignen. (2014, March) Five reasons why feedback may be the most important skill. [Online]. Available: http://www.cambridge.org/elt/blog/2014/03/17/five-reasons-feedback-may-important-skill/

[2] A. Hevner and S. Chatterjee, *Design research in information systems: theory and practice.* New York: Springer Science and Business Media, 2010, ch. Chapter 2.

[3] S. A. Rompf, *Trust and Rationality. An Integrative Framework for Trust Research.* Springer Fachmedien Wiesbaden, 2015, ch. Chapter 2.

[4] H. Yu, Z. Shen, C. Miao, and B. An, "Challenges and opportunities for trust management in crowdsourcing," in *Proceedings - 2012 IEEE/WIC/ACM International Conference on Intelligent Agent Technology, IAT 2012*, vol. 2, 12 2012, pp. 486–493.

[5] J. Howe. (2006, June) Crowdsourcing: A definition. crowdsourcing blog. [Online]. Available: http://crowdsourcing.typepad.com/cs/2006/06/crowdsourcing_a.html

[6] D. Geiger, E. Fielt, M. Rosemann, and M. Schader. (2012) Crowdsourcing information retrieval - definition, typology, and design. [Online]. Available: https://pdfs.semanticscholar.org/e8c6/f13256b9423c67c59599b283cca868a5649e.pdf

[7] F. Olusegun and A. M. Olusegun, "A fuzzy expert system to trust-based access control in crowdsourcing environment," pp. 116–129, 2014. [Online]. Available: http://dx.doi.org/10.1016/j.aci.2014.07.001

[8] G. Anders, H. Seebach, J.-P. Steghöfer, W. Reif, E. André, J. Hähner, C. Müller-Schloer, and T. Ungerer, *The Social Concept of Trust as Enabler for Robustness in Open Self-Organizing Systems.*, 2016, ch. Chapter 1.

[9] D. H. McKnight and N. L. Chervany, "The meaning of trust." Tech. Rep., 1996. [Online]. Available: http://www.misrc.umn.edu/workingpapers/fullpapers/1996/9604_040100.pdf

[10] D. Gambetta, "Can we trust, trust?" pp. 213–237, 2000. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.24.5695&rep=rep1&type=pdf1

[11] A. Jøsang, C. Keser, and T. Dimitrakos, "Can we manage trust?" Springer, Berlin, Heidelberg, 2005. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.60.6731&rep=rep1&type=pdf

[12] A. R. Hevner, S. T. March, J. Park, and S. Ram, "Design science in information system research," pp. 75–105, 2004. [Online]. Available: http://wise.vub.ac.be/sites/default/files/thesis_info/design_science.pdf

[13] S. A. Rompf, *Trust and Rationality. An Integrative Framework for Trust Research.* Springer Fachmedien Wiesbaden, 2015, ch. Chapter 1.

[14] K. Normandeu. (2013) Beyond volume, variety and velocity is the issue of big data veracity. [Online]. Available: https://insidebigdata.com/2013/09/12/beyond-volume-variety-velocity-issue-big-data-veracity/

[15] M. Walker. (2012) Data veracity. [Online]. Available: https://www.datasciencecentral.com/profiles/blogs/data-veracity

[16] E. Chang, P. Thomson, T. Dillon, and F. Hussain, "The fuzzy and dynamic nature of trust," *Trust, Privacy, and Security in Digital Business. TrustBus 2005. Lecture Notes in Computer Science*, vol. 3592, 2005. [Online]. Available: https://pdfs.semanticscholar.org/c838/8ab52c27e5da23fbf8501f7a8f8d0321cfae.pdf

[17] F. Addicts. (2017) Football addicts. [Online]. Available: https://www.footballaddicts.com/apps/

[18] N. Luhmann, *Trust: Making and Breaking Cooperative Relations.* Department of Sociology, University of Oxford, 1980, ch. Familiarity, Confidence, Trust: Problems and Alternatives, pp. 94–105. [Online]. Available: https://www.nuffield.ox.ac.uk/users/gambetta/Trust_making%20and%20breaking%20cooperative%20relations.pdf

[19] J. Jalava, "Trust as a decision," 2006. [Online]. Available: http://ethesis.helsinki.fi/julkaisut/val/sospo/vk/jalava/trustasa.pdf

[20] M. Pavlidis, "Designing for trust." pp. 1–9, 2011. [Online]. Available: http://ceur-ws.org/Vol-731/03.pdf

[21] D. Lekkas, "Establishing and managing trust within the public key infrastructure," *Comput. Commun.*, vol. 26, no. 16, pp. 1815–1825, Oct. 2003. [Online]. Available: https://pdfs.semanticscholar.org/7510/882a693cf432ddc694c9490d2f980ade53e4.pdf

[22] S. A. Rompf, *Trust and Rationality. An Integrative Framework for Trust Research.* Springer Fachmedien Wiesbaden, 2015.

[23] Y. Wang and M. P. Singh, "Formal trust model for multiagent systems," in *Proceedings of the 20th International Joint Conference on Artifical Intelligence*, ser. IJCAI'07. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2007, pp. 1551–1556. [Online]. Available: https://www.ijcai.org/Proceedings/07/Papers/250.pdf

[24] N. Luhmann, *Trust and Power.* Chichester [etc.] : Wiley, 197, 1979.

[25] Y. Han, "Building robust crowdsourcing systems with reputation-aware decision support techniques." 2015. [Online]. Available: https://arxiv.org/pdf/1502.02106.pdf

[26] Trustpilot. (2017) Trustpilot is an online review community. [Online]. Available: https://www.trustpilot.com/about

[27] S. D. Kamvar, M. T. Schlosser, and H. Garcia-Molina., "The eigentrust algorithm for reputation management in p2p networks." 2003. [Online]. Available: http://ilpubs.stanford.edu:8090/562/1/2002-56.pdf

[28] A. Jøsang and R. Ismail, "The beta reputation system," in *In Proceedings of the 15th Bled Conference on Electronic Commerce*, 2002.

[29] C. Castelfranchi and R. Falcone, "2005. socio-cognitive theory of trust," in *In Open Agent Societies: Normative Specifications in Multi-Agent*. Wiley, 2005.

[30] R. Falcone and C. CastelFranchi. (2009) Socio-cognitive model of trust. [Online]. Available: http://ebooks.narotama.ac.id/files/Encyclopedia%20of%20Information%20Science%20and%20Technology%20(2nd%20Edition)/Socio-Cognitive%20Model%20of%20Trust.pdf

[31] M. J. Kollingbaum, "Supervised interaction – creating a web of trust for contracting agents in electronic environments," in *In Proceedings of the First International Joint Conference on Autonomous Agents and Multi–Agent Systems*. ACM Press, 2002, pp. 272–279.

[32] B. Yu and M. P. Sing, "Searching social networks," in *In: Proceedings of the 2nd International Joint Conference on Autonomous Agent and Multi-Agent Systems (AAMAS)*. ACM Press, New York, 2003. [Online]. Available: http://www.divms.uiowa.edu/~ghosh/p65-yu.pdf

[33] J. Weng, C. Miao, A. Goh, Z. Shen, and R. Gay, "Trust-based agent community for collaborative recommendation," in *Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems*, ser. AAMAS '06. New York, NY, USA: ACM, 2006, pp. 1260–1262. [Online]. Available: http://doi.acm.org/10.1145/1160633.1160860

[34] V. Muñoz, J. Murillo, B. López, and D. Busquets, *Strategies for Exploiting Trust Models in Competitive Multi-Agent Systems*. Springer, Berling, Heidelberg, 2009, vol. 5774.

[35] J. Howe. (2006, June) The rise of crowdsourcing. [Online]. Available: https://www.wired.com/2006/06/crowds/

[36] Y. Benkler, "Peer production and cooperation, forthcoming in j. m. bauer m. latzer (eds.)," *Handbook on the Economics of the Internet, Cheltenham and Northampton, Edward Elgar.*, 2016. [Online]. Available: http://www.benkler.org/Peer%20production%20and%20cooperation%2009.pdf

[37] J. Feller, P. F. P, J. Hayes, and P. O'Reilly., "Orchestrating'sustainable crowdsourcing: A characterisation of solver brokerages." *The Journal of Strategic Information Systems*, pp. 216–232, 2012. [Online]. Available: http://www.benkler.org/Peer%20production%20and%20cooperation%2009.pdf

[38] D. C. Brabham, *Crowdsourcing*. Springer, 2013.

[39] E. Estellés-Arolas and F. González-Ladrón-De-Guevara, "Towards an integrated crowdsourcing definition," *Journal of Information Science*, vol. 38, no. 2, pp. 189–200, Apr. 2012. [Online]. Available: http://dx.doi.org/10.1177/0165551512437638

[40] K. Dunbar, "Problem solving," in *A companion to Cognitive Science*, 1998, pp. 289–298. [Online]. Available: http://www.utsc.utoronto.ca/~dunbarlab/pubpdfs/Dunbar%20-%20Problem%20Solving.pdf

[41] E. Martineau, "A typology of crowdsourcing participation styles," April 2012. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.632.9955&rep=rep1&type=pdf

[42] J. Howe, "Crowdsourcing. why the power of the crowd is driving the future of business." 2008. [Online]. Available: http://www.bizbriefings.com/Samples/IntInst%20---%20Crowdsourcing.PDF

[43] A. Wiggins and K. Crowston, "From conservation to crowdsourcing: A typology of citizen science," 2011. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.194.920&rep=rep1&type=pdf

[44] R. M. Ryan and E. L. Deci, "Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being," 2000. [Online]. Available: https://selfdeterminationtheory.org/SDT/documents/2000_RyanDeci_SDT.pdf

[45] M. Venanzi, A. Rogers, and N. R. Jennings, "Crowdsourcing spatial phenomena using trust-based heteroskedastic gaussian processes," in *HCOMP*. AAAI, 2013.

[46] G. Riccardi, A. Ghosh, S. Chowdhury, and A. Orkan Bayer, "Motivational feedback in crowdsourcing: a case study in speech transcription," in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 08 2013.

[47] J. Hattie and H. Timperley, "The power of feedback," *Review of Educational Research*, vol. 77, pp. 81–112, 2007. [Online]. Available: http://www.columbia.edu/~mvp19/ETF/Feedback.pdf

[48] D. L. Butler and P. Winne, "Feedback and self-regulated learning: A theoretical synthesis," vol. 65, pp. 245–281, 09 1995.

[49] S. M. Brookhart, *How To Give Effective Feedback To Your Students*. Association for Supervision and Curriculum Development, 2008.

[50] Sportswik. (2017) Så här fungerar sportswik. [Online]. Available: https://www.sportswik.com/Content/Sportswik_infoblad_web.pdf

[51] ——. (2017) Liverapportera från matcher. [Online]. Available: https://www.sportswik.com/

[52] Crowdscores. (2017) What we do. [Online]. Available: http://crowdscores.com/about-us/

[53] GoalShouter. (2018) Description. [Online]. Available: https://itunes.apple.com/app/goalshouter/id428932802

[54] ——. (2018) How does it work. [Online]. Available: https://www.goalshouter.com/

[55] ScoreStream. (2018) About. [Online]. Available: https://scorestream.com/about

[56] ——. (2018) How it works. [Online]. Available: https://scorestream.com/partners

[57] W. Foundation. (2017). [Online]. Available: https://wikimediafoundation.org/wiki/Home

[58] ——. (2017) Our projects. [Online]. Available: https://wikimediafoundation.org/wiki/Our_projects

[59] eBay. (2017) How feedback works. [Online]. Available: http://pages.ebay.com/help/feedback/howitworks.html

[60] P. Resnick, R. Zeckhauser, and J. Swanson, "The value of reputation on ebay: a controlled experiment." pp. 1–47, 2006. [Online]. Available: http://presnick.people.si.umich.edu/papers/postcards/PostcardsFinalPrePub.pdf

[61] TripAdvisor. (2017) How does the popularity index differ from the overall bubble rating? [Online]. Avail-

able: https://www.tripadvisorsupport.com/hc/en-us/articles/200614027-How-does-the-Popularity-Index-differ-from-the-overall-bubble-rating-

[62] ——. (2017) Everything you need to know about the tripadvisor popularity ranking algorithm. [Online]. Available: https://www.tripadvisor.ca/TripAdvisorInsights/n2701/everything-you-need-know-about-tripadvisor-popularity-ranking-algorithm

[63] ——. (2017) What is a rating? [Online]. Available: https://www.tripadvisorsupport.com/hc/en-us/articles/200613867-What-is-a-rating-

[64] J. W. Creswell, *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, 3rd ed. Sage Publications Ltd., 2008.

A

Material used in the validation sessions based on users history of reported lineups

The following material was used in the test session that took place in the **first** iteration. Note that the examples of a users history of reported lineups, corresponding to a certain trust value, is shown in an increased order of trust values, and not in the order in which they were shown to the participants in the test sessions.

		Date of report	User id	Approved	Discarded		
210	274	2017-09-25 15:55:01 UTC	8DC09A93-94F3-4F15-85D6-67D98C216280	TRUE	FALSE		
211	275	2017-09-25 15:56:22 UTC	8DC09A93-94F3-4F15-85D6-67D98C216280	TRUE	FALSE		
212	276	2017-09-25 15:57:52 UTC	8DC09A93-94F3-4F15-85D6-67D98C216280	TRUE	FALSE		
213	277	2017-09-25 15:59:20 UTC	8DC09A93-94F3-4F15-85D6-67D98C216280	TRUE	FALSE		
214	278	2017-09-25 16:01:56 UTC	8DC09A93-94F3-4F15-85D6-67D98C216280	TRUE	FALSE		
215	279	2017-09-25 16:03:37 UTC	8DC09A93-94F3-4F15-85D6-67D98C216280	TRUE	FALSE		
216	280	2017-09-25 16:05:36 UTC	8DC09A93-94F3-4F15-85D6-67D98C216280	TRUE	FALSE		
217	281	2017-09-25 16:06:40 UTC	8DC09A93-94F3-4F15-85D6-67D98C216280	TRUE	FALSE		
218	282	2017-09-25 17:57:36 UTC	8DC09A93-94F3-4F15-85D6-67D98C216280	TRUE	FALSE		
219	283	2017-09-25 18:00:45 UTC	8DC09A93-94F3-4F15-85D6-67D98C216280	TRUE	FALSE		
220	284	2017-09-26 12:45:33 UTC	7F9C4A96-440F-498E-A8A8-FBF5645467FB	FALSE	TRUE		
221	285	2017-09-26 14:52:01 UTC	0597831D-D00B-434C-A89A-AE0780A4857B	FALSE	TRUE		
222	286	2017-09-26 16:40:53 UTC	25F13E57-F839-4955-8BC8-EFC86585AEE4	FALSE	TRUE		
223	287	2017-09-26 17:16:49 UTC	9563EA97-8BAF-4157-A98A-0A2062970184	FALSE	TRUE		
224	288	2017-09-26 17:24:35 UTC	23875E9E-F787-4E7F-A4EA-03B60497AE8E	FALSE	TRUE		
225	289	2017-09-26 17:26:43 UTC	23875E9E-F787-4E7F-A4EA-03B60497AE8E	FALSE	TRUE		
226	290	2017-09-26 18:01:15 UTC	1DE6FD35-3A98-40AC-AEDB-80C12420FF7E	FALSE	FALSE		
227	291	2017-09-26 18:07:43 UTC	1DE6FD35-3A98-40AC-AEDB-80C12420FF7E	FALSE	FALSE		
228	292	2017-09-26 18:10:38 UTC	1DE6FD35-3A98-40AC-AEDB-80C12420FF7E	FALSE	FALSE		
229	293	2017-09-26 18:11:57 UTC	F215719A-631B-414F-AA5D-7CCC8B43CCED	FALSE	TRUE		
230	294	2017-09-26 18:12:42 UTC	1DE6FD35-3A98-40AC-AEDB-80C12420FF7E	FALSE	FALSE		
231	295	2017-09-26 18:13:18 UTC	F215719A-631B-414F-AA5D-7CCC8B43CCED	FALSE	FALSE		
232	296	2017-09-26 19:22:03 UTC	69ACA14E-8025-42A5-9720-03053D9A73E1	FALSE	FALSE		
233	297	2017-09-26 20:34:22 UTC	0597831D-D00B-434C-A89A-AE0780A4857B	FALSE	FALSE		
234	298	2017-09-26 20:37:33 UTC	B4ACA71A-CECF-4EF3-BF9E-B4C45CE1D1BB	FALSE	FALSE		
235	299	2017-09-26 21:53:17 UTC	22D00AE8-F918-4897-B05F-7AA337CEA15E	FALSE	FALSE		
236	300	2017-09-27 05:46:43 UTC	D64A537C-C657-4C25-906C-D61BBAFAC7D5	FALSE	FALSE		
237	301	2017-09-27 05:50:00 UTC	D64A537C-C657-4C25-906C-D61BBAFAC7D5	FALSE	FALSE		
238	302	2017-09-27 06:16:06 UTC	BD6AF598-9659-407F-9687-9EE8B80FCA32	FALSE	FALSE		
239	303	2017-09-27 06:52:03 UTC	69ACA14E-8025-42A5-9720-03053D9A73E1	FALSE	FALSE		

Figure A.1: An example of users that have reported lineups under a certain timespan.

Reported At	Reported At (Day)	Decision	Reporter Forza ID
Dec 10, 2017 16:10	December 10, 2017	APPROVED	ADBB81AB-CD73-4EE6-AF57-324837F21BEC
Dec 17, 2017 10:45	December 17, 2017	APPROVED	ADBB81AB-CD73-4EE6-AF57-324837F21BEC
Dec 17, 2017 10:47	December 17, 2017	APPROVED	ADBB81AB-CD73-4EE6-AF57-324837F21BEC
Dec 23, 2017 12:58	December 23, 2017	DISCARDED	ADBB81AB-CD73-4EE6-AF57-324837F21BEC
Dec 23, 2017 12:59	December 23, 2017	DISCARDED	ADBB81AB-CD73-4EE6-AF57-324837F21BEC

Figure A.2: An example of a users history of reported lineups corresponding to a trust value of 0.

Reported At	Reported At (Day)	Decision	Reporter Forza ID
Sep 26, 2017 18:01	September 26, 2017	APPROVED	1DE6FD35-3A98-40AC-AEDB-80C12420FF7E
Sep 26, 2017 18:07	September 26, 2017	DISCARDED	1DE6FD35-3A98-40AC-AEDB-80C12420FF7E
Sep 26, 2017 18:10	September 26, 2017	DISCARDED	1DE6FD35-3A98-40AC-AEDB-80C12420FF7E
Sep 26, 2017 18:12	September 26, 2017	DISCARDED	1DE6FD35-3A98-40AC-AEDB-80C12420FF7E
Sep 27, 2017 18:22	September 27, 2017	DISCARDED	1DE6FD35-3A98-40AC-AEDB-80C12420FF7E
Sep 27, 2017 18:24	September 27, 2017	DISCARDED	1DE6FD35-3A98-40AC-AEDB-80C12420FF7E
Oct 01, 2017 12:01	October 1, 2017	APPROVED	1DE6FD35-3A98-40AC-AEDB-80C12420FF7E
Mar 27, 2018 17:20	March 27, 2018	APPROVED	1DE6FD35-3A98-40AC-AEDB-80C12420FF7E
Mar 27, 2018 17:27	March 27, 2018	APPROVED	1DE6FD35-3A98-40AC-AEDB-80C12420FF7E
Mar 27, 2018 18:05	March 27, 2018	DISCARDED	1DE6FD35-3A98-40AC-AEDB-80C12420FF7E

Figure A.3: An example of a users history of reported lineups corresponding to a trust value of 0.1.

Reported At	Reported At (Day)	Decision	Reporter Forza ID
Mar 04, 2018 12:57	March 4, 2018	APPROVED	AA53774B-6A54-428F-95DB-395DE2871253
Mar 24, 2018 14:49	March 24, 2018	DISCARDED	AA53774B-6A54-428F-95DB-395DE2871253
Mar 24, 2018 15:21	March 24, 2018	APPROVED	AA53774B-6A54-428F-95DB-395DE2871253
Mar 29, 2018 13:23	March 29, 2018	APPROVED	AA53774B-6A54-428F-95DB-395DE2871253

Figure A.4: An example of a users history of reported lineups corresponding to a trust value of 0.2.

Reported At	Reported At (Day)	Decision	Reporter Forza ID
Jan 27, 2018 16:04	January 27, 2018	APPROVED	2292CACA-B628-4D70-96FA-0E7E8704CE03
Jan 27, 2018 18:28	January 27, 2018	DISCARDED	2292CACA-B628-4D70-96FA-0E7E8704CE03
Jan 27, 2018 18:44	January 27, 2018	APPROVED	2292CACA-B628-4D70-96FA-0E7E8704CE03
Jan 28, 2018 18:40	January 28, 2018	DISCARDED	2292CACA-B628-4D70-96FA-0E7E8704CE03
Feb 04, 2018 10:16	February 4, 2018	DISCARDED	2292CACA-B628-4D70-96FA-0E7E8704CE03
Feb 04, 2018 10:19	February 4, 2018	DISCARDED	2292CACA-B628-4D70-96FA-0E7E8704CE03
Feb 04, 2018 12:54	February 4, 2018	DISCARDED	2292CACA-B628-4D70-96FA-0E7E8704CE03
Feb 04, 2018 12:56	February 4, 2018	APPROVED	2292CACA-B628-4D70-96FA-0E7E8704CE03
Feb 04, 2018 13:00	February 4, 2018	DISCARDED	2292CACA-B628-4D70-96FA-0E7E8704CE03
Feb 04, 2018 13:01	February 4, 2018	DISCARDED	2292CACA-B628-4D70-96FA-0E7E8704CE03
Feb 09, 2018 18:28	February 9, 2018	DISCARDED	2292CACA-B628-4D70-96FA-0E7E8704CE03
Feb 09, 2018 18:29	February 9, 2018	DISCARDED	2292CACA-B628-4D70-96FA-0E7E8704CE03
Feb 17, 2018 16:05	February 17, 2018	APPROVED	2292CACA-B628-4D70-96FA-0E7E8704CE03
Feb 18, 2018 10:18	February 18, 2018	DISCARDED	2292CACA-B628-4D70-96FA-0E7E8704CE03
Feb 18, 2018 10:20	February 18, 2018	DISCARDED	2292CACA-B628-4D70-96FA-0E7E8704CE03
Feb 25, 2018 11:02	February 25, 2018	APPROVED	2292CACA-B628-4D70-96FA-0E7E8704CE03
Feb 26, 2018 18:36	February 26, 2018	DISCARDED	2292CACA-B628-4D70-96FA-0E7E8704CE03
Feb 26, 2018 18:38	February 26, 2018	APPROVED	2292CACA-B628-4D70-96FA-0E7E8704CE03
Mar 09, 2018 18:28	March 9, 2018	DISCARDED	2292CACA-B628-4D70-96FA-0E7E8704CE03
Mar 09, 2018 18:30	March 9, 2018	APPROVED	2292CACA-B628-4D70-96FA-0E7E8704CE03
Mar 11, 2018 12:56	March 11, 2018	APPROVED	2292CACA-B628-4D70-96FA-0E7E8704CE03
Mar 11, 2018 12:58	March 11, 2018	APPROVED	2292CACA-B628-4D70-96FA-0E7E8704CE03
Mar 11, 2018 12:59	March 11, 2018	DISCARDED	2292CACA-B628-4D70-96FA-0E7E8704CE03
Mar 11, 2018 13:01	March 11, 2018	APPROVED	2292CACA-B628-4D70-96FA-0E7E8704CE03
Mar 18, 2018 10:24	March 18, 2018	APPROVED	2292CACA-B628-4D70-96FA-0E7E8704CE03

Figure A.5: An example of a users history of reported lineups corresponding to a trust value of 0.3.

Reported At	Reported At (Day)	Decision	Reporter Forza ID
Dec 04, 2017 18:28	December 4, 2017	DISCARDED	B4D175C0-DD52-41F9-9E96-44255271F9C3
Dec 04, 2017 18:36	December 4, 2017	DISCARDED	B4D175C0-DD52-41F9-9E96-44255271F9C3
Dec 05, 2017 11:17	December 5, 2017	DISCARDED	B4D175C0-DD52-41F9-9E96-44255271F9C3
Dec 05, 2017 11:21	December 5, 2017	DISCARDED	B4D175C0-DD52-41F9-9E96-44255271F9C3
Dec 05, 2017 11:25	December 5, 2017	DISCARDED	B4D175C0-DD52-41F9-9E96-44255271F9C3
Dec 05, 2017 11:31	December 5, 2017	DISCARDED	B4D175C0-DD52-41F9-9E96-44255271F9C3
Dec 05, 2017 11:34	December 5, 2017	DISCARDED	B4D175C0-DD52-41F9-9E96-44255271F9C3
Dec 06, 2017 18:40	December 6, 2017	APPROVED	B4D175C0-DD52-41F9-9E96-44255271F9C3
Dec 06, 2017 18:44	December 6, 2017	APPROVED	B4D175C0-DD52-41F9-9E96-44255271F9C3
Dec 07, 2017 18:05	December 7, 2017	APPROVED	B4D175C0-DD52-41F9-9E96-44255271F9C3
Dec 09, 2017 18:45	December 9, 2017	APPROVED	B4D175C0-DD52-41F9-9E96-44255271F9C3
Dec 09, 2017 18:47	December 9, 2017	APPROVED	B4D175C0-DD52-41F9-9E96-44255271F9C3
Dec 10, 2017 12:53	December 10, 2017	APPROVED	B4D175C0-DD52-41F9-9E96-44255271F9C3
Dec 10, 2017 12:55	December 10, 2017	DISCARDED	B4D175C0-DD52-41F9-9E96-44255271F9C3
Dec 10, 2017 12:56	December 10, 2017	APPROVED	B4D175C0-DD52-41F9-9E96-44255271F9C3
Dec 10, 2017 18:45	December 10, 2017	DISCARDED	B4D175C0-DD52-41F9-9E96-44255271F9C3
Dec 11, 2017 18:49	December 11, 2017	APPROVED	B4D175C0-DD52-41F9-9E96-44255271F9C3
Dec 12, 2017 18:32	December 12, 2017	APPROVED	B4D175C0-DD52-41F9-9E96-44255271F9C3
Dec 12, 2017 18:57	December 12, 2017	APPROVED	B4D175C0-DD52-41F9-9E96-44255271F9C3
Dec 12, 2017 19:30	December 12, 2017	APPROVED	B4D175C0-DD52-41F9-9E96-44255271F9C3
Dec 13, 2017 18:40	December 13, 2017	DISCARDED	B4D175C0-DD52-41F9-9E96-44255271F9C3
Dec 13, 2017 18:44	December 13, 2017	DISCARDED	B4D175C0-DD52-41F9-9E96-44255271F9C3
Dec 13, 2017 18:48	December 13, 2017	APPROVED	B4D175C0-DD52-41F9-9E96-44255271F9C3

Figure A.6: An example of a users history of reported lineups corresponding to a trust value of 0.4.

Reported At	Reported At (Day)	Decision	Reporter Forza ID
Mar 23, 2018 17:10	March 23, 2018	APPROVED	9480986F-776E-4250-8182-C73D0464FE4A
Mar 23, 2018 17:34	March 23, 2018	APPROVED	9480986F-776E-4250-8182-C73D0464FE4A
Mar 23, 2018 17:38	March 23, 2018	APPROVED	9480986F-776E-4250-8182-C73D0464FE4A
Mar 23, 2018 18:16	March 23, 2018	DISCARDED	9480986F-776E-4250-8182-C73D0464FE4A
Mar 23, 2018 18:17	March 23, 2018	DISCARDED	9480986F-776E-4250-8182-C73D0464FE4A
Mar 23, 2018 18:26	March 23, 2018	APPROVED	9480986F-776E-4250-8182-C73D0464FE4A
Mar 23, 2018 18:28	March 23, 2018	APPROVED	9480986F-776E-4250-8182-C73D0464FE4A
Mar 27, 2018 16:36	March 27, 2018	APPROVED	9480986F-776E-4250-8182-C73D0464FE4A
Mar 27, 2018 16:38	March 27, 2018	APPROVED	9480986F-776E-4250-8182-C73D0464FE4A
Mar 27, 2018 18:49	March 27, 2018	APPROVED	9480986F-776E-4250-8182-C73D0464FE4A

Figure A.7: An example of a users history of reported lineups corresponding to a trust value of 0.5.

Reported At	Reported At (Day)	Decision	Reporter Forza ID
Dec 18, 2017 20:59	December 18, 2017	APPROVED	E4CBFCF8-E500-47AD-A2E9-3A0E675105E4
Jan 28, 2018 13:08	January 28, 2018	APPROVED	E4CBFCF8-E500-47AD-A2E9-3A0E675105E4
Jan 28, 2018 13:31	January 28, 2018	APPROVED	E4CBFCF8-E500-47AD-A2E9-3A0E675105E4
Mar 20, 2018 20:01	March 20, 2018	APPROVED	E4CBFCF8-E500-47AD-A2E9-3A0E675105E4
Mar 29, 2018 17:56	March 29, 2018	APPROVED	E4CBFCF8-E500-47AD-A2E9-3A0E675105E4

Figure A.8: An example of a users history of reported lineups corresponding to a trust value of 0.6.

Reported At	Reported At (Day)	Decision	Reporter Forza ID
Oct 28, 2017 13:38	October 28, 2017	DISCARDED	BDDD8052-75DA-491E-9CEC-82D73F661213
Oct 28, 2017 13:40	October 28, 2017	APPROVED	BDDD8052-75DA-491E-9CEC-82D73F661213
Oct 28, 2017 13:43	October 28, 2017	APPROVED	BDDD8052-75DA-491E-9CEC-82D73F661213
Oct 28, 2017 13:44	October 28, 2017	APPROVED	BDDD8052-75DA-491E-9CEC-82D73F661213
Oct 28, 2017 13:52	October 28, 2017	APPROVED	BDDD8052-75DA-491E-9CEC-82D73F661213
Oct 28, 2017 14:24	October 28, 2017	APPROVED	BDDD8052-75DA-491E-9CEC-82D73F661213
Oct 28, 2017 14:28	October 28, 2017	APPROVED	BDDD8052-75DA-491E-9CEC-82D73F661213
Oct 28, 2017 14:31	October 28, 2017	APPROVED	BDDD8052-75DA-491E-9CEC-82D73F661213

Figure A.9: An example of a users history of reported lineups corresponding to a trust value of 0.7.

Reported At	Reported At (Day)	Decision	Reporter Forza ID
Mar 31, 2018 21:41	March 31, 2018	DISCARDED	53F9A876-37ED-42F1-8933-86C6B764CC99
Mar 22, 2018 22:08	March 22, 2018	APPROVED	53F9A876-37ED-42F1-8933-86C6B764CC99
Mar 04, 2018 16:54	March 4, 2018	APPROVED	53F9A876-37ED-42F1-8933-86C6B764CC99
Feb 28, 2018 18:26	February 28, 2018	APPROVED	53F9A876-37ED-42F1-8933-86C6B764CC99
Feb 24, 2018 14:01	February 24, 2018	APPROVED	53F9A876-37ED-42F1-8933-86C6B764CC99
Feb 12, 2018 18:47	February 12, 2018	APPROVED	53F9A876-37ED-42F1-8933-86C6B764CC99
Feb 11, 2018 19:59	February 11, 2018	APPROVED	53F9A876-37ED-42F1-8933-86C6B764CC99
Feb 01, 2018 0:41	February 1, 2018	DISCARDED	53F9A876-37ED-42F1-8933-86C6B764CC99
Jan 31, 2018 18:52	January 31, 2018	APPROVED	53F9A876-37ED-42F1-8933-86C6B764CC99
Jan 07, 2018 9:57	January 7, 2018	APPROVED	53F9A876-37ED-42F1-8933-86C6B764CC99
Jan 06, 2018 18:48	January 6, 2018	APPROVED	53F9A876-37ED-42F1-8933-86C6B764CC99
Dec 29, 2017 11:13	December 29, 2017	APPROVED	53F9A876-37ED-42F1-8933-86C6B764CC99
Dec 28, 2017 19:07	December 28, 2017	APPROVED	53F9A876-37ED-42F1-8933-86C6B764CC99
Dec 05, 2017 17:57	December 5, 2017	APPROVED	53F9A876-37ED-42F1-8933-86C6B764CC99
Dec 03, 2017 10:30	December 3, 2017	DISCARDED	53F9A876-37ED-42F1-8933-86C6B764CC99

Figure A.10: An example of a users history of reported lineups corresponding to a trust value of 0.8.

Reported At	Reported At (Day)	Decision	Reporter Forza ID
Dec 17, 2017 15:15	December 17, 2017	APPROVED	032728A0-A000-42B1-8B4C-D5659AADA78A
Jan 28, 2018 16:18	January 28, 2018	APPROVED	032728A0-A000-42B1-8B4C-D5659AADA78A
Jan 28, 2018 16:23	January 28, 2018	APPROVED	032728A0-A000-42B1-8B4C-D5659AADA78A
Feb 04, 2018 13:07	February 4, 2018	DISCARDED	032728A0-A000-42B1-8B4C-D5659AADA78A
Feb 04, 2018 13:23	February 4, 2018	APPROVED	032728A0-A000-42B1-8B4C-D5659AADA78A
Feb 09, 2018 19:57	February 9, 2018	APPROVED	032728A0-A000-42B1-8B4C-D5659AADA78A
Feb 18, 2018 15:04	February 18, 2018	APPROVED	032728A0-A000-42B1-8B4C-D5659AADA78A
Feb 18, 2018 15:09	February 18, 2018	APPROVED	032728A0-A000-42B1-8B4C-D5659AADA78A
Feb 25, 2018 15:51	February 25, 2018	APPROVED	032728A0-A000-42B1-8B4C-D5659AADA78A
Feb 25, 2018 15:55	February 25, 2018	APPROVED	032728A0-A000-42B1-8B4C-D5659AADA78A
Feb 25, 2018 15:56	February 25, 2018	APPROVED	032728A0-A000-42B1-8B4C-D5659AADA78A

Figure A.11: An example of a users history of reported lineups corresponding to a trust value of 0.9.

Reported At	Reported At (Day)	Decision	Reporter Forza ID
Sep 26, 2017 21:53	September 26, 2017	APPROVED	22D00AE8-F918-4897-B05F-7AA337CEA15E
Oct 14, 2017 13:07	October 14, 2017	APPROVED	22D00AE8-F918-4897-B05F-7AA337CEA15E
Nov 22, 2017 19:40	November 22, 2017	APPROVED	22D00AE8-F918-4897-B05F-7AA337CEA15E
Dec 01, 2017 18:57	December 1, 2017	APPROVED	22D00AE8-F918-4897-B05F-7AA337CEA15E
Dec 09, 2017 14:21	December 9, 2017	APPROVED	22D00AE8-F918-4897-B05F-7AA337CEA15E
Dec 16, 2017 14:02	December 16, 2017	APPROVED	22D00AE8-F918-4897-B05F-7AA337CEA15E
Jan 07, 2018 11:26	January 7, 2018	APPROVED	22D00AE8-F918-4897-B05F-7AA337CEA15E
Jan 13, 2018 14:11	January 13, 2018	APPROVED	22D00AE8-F918-4897-B05F-7AA337CEA15E
Jan 20, 2018 14:17	January 20, 2018	APPROVED	22D00AE8-F918-4897-B05F-7AA337CEA15E
Jan 30, 2018 18:51	January 30, 2018	DISCARDED	22D00AE8-F918-4897-B05F-7AA337CEA15E
Feb 18, 2018 15:36	February 18, 2018	APPROVED	22D00AE8-F918-4897-B05F-7AA337CEA15E
Feb 24, 2018 14:08	February 24, 2018	APPROVED	22D00AE8-F918-4897-B05F-7AA337CEA15E

Figure A.12: An example of a users history of reported lineups corresponding to a trust value of 1.0.

The following material was used in the test session that took place in the **second** iteration. Note that the examples of a users history of reported lineups, corresponding to a certain trust value, is shown in an increased order of trust values, and not in the order in which they were shown to the participants in the test sessions.

Reported At	Decision	Reporter Forza ID	Quality
Sep 27, 2017 17:42	DISCARDED	050BCA8D-8FC5-494A-ABAE-2467C050ABE1	
Sep 30, 2017 12:24	DISCARDED	050BCA8D-8FC5-494A-ABAE-2467C050ABE1	
Sep 30, 2017 12:26	DISCARDED	050BCA8D-8FC5-494A-ABAE-2467C050ABE1	
Oct 07, 2017 6:29	APPROVED	050BCA8D-8FC5-494A-ABAE-2467C050ABE1	Approved with one adjustment
Oct 28, 2017 10:37	APPROVED	050BCA8D-8FC5-494A-ABAE-2467C050ABE1	Approved with three adjustments

Figure A.13: An example of a users history of reported lineups corresponding to a trust value of 0.06.

Reported At	Decision	Reporter Forza ID	Quality
Oct 10, 2017 15:55	DISCARDED	17BE7B8E-0D65-4AB4-8970-237EED6C54C4	
Oct 10, 2017 15:57	DISCARDED	17BE7B8E-0D65-4AB4-8970-237EED6C54C4	
Nov 08, 2017 19:02	APPROVED	17BE7B8E-0D65-4AB4-8970-237EED6C54C4	
Mar 17, 2018 18:33	DISCARDED	17BE7B8E-0D65-4AB4-8970-237EED6C54C4	
Mar 27, 2018 17:01	APPROVED	17BE7B8E-0D65-4AB4-8970-237EED6C54C4	Approved with two adjustments
Apr 03, 2018 16:07	APPROVED	17BE7B8E-0D65-4AB4-8970-237EED6C54C4	

Figure A.14: An example of a users history of reported lineups corresponding to a trust value of 0.17.

Reported At	Decision	Reporter Forza ID	Quality
Dec 04, 2017 17:06	APPROVED	4FA78021-05C5-4649-89E4-2788BBF61E50	
Dec 04, 2017 17:07	APPROVED	4FA78021-05C5-4649-89E4-2788BBF61E50	
Feb 15, 2018 18:53	DISCARDED	4FA78021-05C5-4649-89E4-2788BBF61E50	
Feb 15, 2018 18:55	APPROVED	4FA78021-05C5-4649-89E4-2788BBF61E50	Approved with two adjustments
Feb 15, 2018 18:58	APPROVED	4FA78021-05C5-4649-89E4-2788BBF61E50	

Figure A.15: An example of a users history of reported lineups corresponding to a trust value of 0.27.

Reported At	Decision	Reporter Forza ID	Quality
Sep 30, 2017 14:59	DISCARDED	1855057F-57E2-4FC5-8FB1-96A9138E25BF	
Dec 28, 2017 19:06	APPROVED	1855057F-57E2-4FC5-8FB1-96A9138E25BF	
Jan 07, 2018 13:42	DISCARDED	1855057F-57E2-4FC5-8FB1-96A9138E25BF	
Jan 27, 2018 14:36	DISCARDED	1855057F-57E2-4FC5-8FB1-96A9138E25BF	
Jan 27, 2018 14:47	DISCARDED	1855057F-57E2-4FC5-8FB1-96A9138E25BF	
Jan 29, 2018 14:04	APPROVED	1855057F-57E2-4FC5-8FB1-96A9138E25BF	
Jan 29, 2018 14:10	DISCARDED	1855057F-57E2-4FC5-8FB1-96A9138E25BF	
Feb 17, 2018 15:52	APPROVED	1855057F-57E2-4FC5-8FB1-96A9138E25BF	Approved with one adjustment
Mar 03, 2018 14:50	APPROVED	1855057F-57E2-4FC5-8FB1-96A9138E25BF	
Mar 24, 2018 11:03	APPROVED	1855057F-57E2-4FC5-8FB1-96A9138E25BF	
Apr 02, 2018 12:13	APPROVED	1855057F-57E2-4FC5-8FB1-96A9138E25BF	Approved with one adjustment
Apr 02, 2018 12:19	APPROVED	1855057F-57E2-4FC5-8FB1-96A9138E25BF	

Figure A.16: An example of a users history of reported lineups corresponding to a trust value of 0.34.

Reported At	Decision	Reporter Forza ID	Quality
Dec 04, 2017 18:28	DISCARDED	B4D175C0-DD52-41F9-9E96-44255271F9C3	
Dec 04, 2017 18:36	DISCARDED	B4D175C0-DD52-41F9-9E96-44255271F9C3	
Dec 05, 2017 11:17	DISCARDED	B4D175C0-DD52-41F9-9E96-44255271F9C3	
Dec 05, 2017 11:21	DISCARDED	B4D175C0-DD52-41F9-9E96-44255271F9C3	
Dec 05, 2017 11:25	DISCARDED	B4D175C0-DD52-41F9-9E96-44255271F9C3	
Dec 05, 2017 11:31	DISCARDED	B4D175C0-DD52-41F9-9E96-44255271F9C3	
Dec 05, 2017 11:34	DISCARDED	B4D175C0-DD52-41F9-9E96-44255271F9C3	
Dec 06, 2017 18:38	APPROVED	B4D175C0-DD52-41F9-9E96-44255271F9C3	
Dec 06, 2017 18:40	APPROVED	B4D175C0-DD52-41F9-9E96-44255271F9C3	
Dec 06, 2017 18:44	APPROVED	B4D175C0-DD52-41F9-9E96-44255271F9C3	
Dec 07, 2017 18:05	APPROVED	B4D175C0-DD52-41F9-9E96-44255271F9C3	
Dec 09, 2017 18:45	APPROVED	B4D175C0-DD52-41F9-9E96-44255271F9C3	
Dec 09, 2017 18:47	APPROVED	B4D175C0-DD52-41F9-9E96-44255271F9C3	
Dec 10, 2017 12:53	APPROVED	B4D175C0-DD52-41F9-9E96-44255271F9C3	Approved with three adjustments
Dec 10, 2017 12:55	DISCARDED	B4D175C0-DD52-41F9-9E96-44255271F9C3	
Dec 10, 2017 12:56	APPROVED	B4D175C0-DD52-41F9-9E96-44255271F9C3	
Dec 10, 2017 18:45	DISCARDED	B4D175C0-DD52-41F9-9E96-44255271F9C3	
Dec 11, 2017 18:49	APPROVED	B4D175C0-DD52-41F9-9E96-44255271F9C3	
Dec 12, 2017 18:32	APPROVED	B4D175C0-DD52-41F9-9E96-44255271F9C3	
Dec 12, 2017 18:57	APPROVED	B4D175C0-DD52-41F9-9E96-44255271F9C3	
Dec 12, 2017 19:30	APPROVED	B4D175C0-DD52-41F9-9E96-44255271F9C3	
Dec 13, 2017 18:40	DISCARDED	B4D175C0-DD52-41F9-9E96-44255271F9C3	
Dec 13, 2017 18:44	DISCARDED	B4D175C0-DD52-41F9-9E96-44255271F9C3	
Dec 13, 2017 18:48	APPROVED	B4D175C0-DD52-41F9-9E96-44255271F9C3	

Figure A.17: An example of a users history of reported lineups corresponding to a trust value of 0.44.

Reported At	Decision	Reporter Forza ID	Quality
Mar 04, 2018 21:26	APPROVED	6E675F72-2883-4FFD-A0C4-58B950D19798	Approved with one adjustment
Mar 04, 2018 21:36	APPROVED	6E675F72-2883-4FFD-A0C4-58B950D19798	
Mar 07, 2018 18:26	APPROVED	6E675F72-2883-4FFD-A0C4-58B950D19798	
Mar 07, 2018 18:31	APPROVED	6E675F72-2883-4FFD-A0C4-58B950D19798	Approved with one adjustment
Mar 17, 2018 20:06	APPROVED	6E675F72-2883-4FFD-A0C4-58B950D19798	
Mar 17, 2018 20:13	APPROVED	6E675F72-2883-4FFD-A0C4-58B950D19798	

Figure A.18: An example of a users history of reported lineups corresponding to a trust value of 0.54.

Reported At	Decision	Reporter Forza ID	Quality
Sep 23, 2017 15:51	DISCARDED	4F8C4482-332F-43E3-8782-F5F5D3163C88	
Sep 27, 2017 12:59	DISCARDED	4F8C4482-332F-43E3-8782-F5F5D3163C88	
Sep 27, 2017 13:01	DISCARDED	4F8C4482-332F-43E3-8782-F5F5D3163C88	
Sep 27, 2017 13:12	DISCARDED	4F8C4482-332F-43E3-8782-F5F5D3163C88	
Sep 27, 2017 14:00	DISCARDED	4F8C4482-332F-43E3-8782-F5F5D3163C88	
Feb 13, 2018 18:33	DISCARDED	4F8C4482-332F-43E3-8782-F5F5D3163C88	
Feb 13, 2018 18:35	DISCARDED	4F8C4482-332F-43E3-8782-F5F5D3163C88	
Feb 25, 2018 15:04	APPROVED	4F8C4482-332F-43E3-8782-F5F5D3163C88	Approved with one adjustment
Mar 01, 2018 16:18	APPROVED	4F8C4482-332F-43E3-8782-F5F5D3163C88	Approved with one adjustment
Mar 01, 2018 16:26	APPROVED	4F8C4482-332F-43E3-8782-F5F5D3163C88	
Mar 04, 2018 15:02	APPROVED	4F8C4482-332F-43E3-8782-F5F5D3163C88	
Mar 04, 2018 15:07	APPROVED	4F8C4482-332F-43E3-8782-F5F5D3163C88	
Mar 06, 2018 18:41	APPROVED	4F8C4482-332F-43E3-8782-F5F5D3163C88	
Mar 17, 2018 14:34	APPROVED	4F8C4482-332F-43E3-8782-F5F5D3163C88	Approved with one adjustment
Mar 17, 2018 14:37	APPROVED	4F8C4482-332F-43E3-8782-F5F5D3163C88	
Mar 17, 2018 14:37	APPROVED	4F8C4482-332F-43E3-8782-F5F5D3163C88	
Mar 23, 2018 20:47	DISCARDED	4F8C4482-332F-43E3-8782-F5F5D3163C88	
Mar 28, 2018 6:22	DISCARDED	4F8C4482-332F-43E3-8782-F5F5D3163C88	
Apr 01, 2018 14:42	APPROVED	4F8C4482-332F-43E3-8782-F5F5D3163C88	
Apr 01, 2018 14:43	APPROVED	4F8C4482-332F-43E3-8782-F5F5D3163C88	
Apr 14, 2018 13:02	APPROVED	4F8C4482-332F-43E3-8782-F5F5D3163C88	
Apr 14, 2018 13:08	APPROVED	4F8C4482-332F-43E3-8782-F5F5D3163C88	

Figure A.19: An example of a users history of reported lineups corresponding to a trust value of 0.66.

Reported At	Decision	Reporter Forza ID	Quality
Feb 13, 2018 18:33	DISCARDED	07A0A120-EC0D-4007-B19E-C50D79AA3611	
Feb 13, 2018 19:36	APPROVED	07A0A120-EC0D-4007-B19E-C50D79AA3611	
Feb 13, 2018 19:42	APPROVED	07A0A120-EC0D-4007-B19E-C50D79AA3611	Approved with two adjustments
Feb 13, 2018 20:48	APPROVED	07A0A120-EC0D-4007-B19E-C50D79AA3611	
Feb 13, 2018 20:53	APPROVED	07A0A120-EC0D-4007-B19E-C50D79AA3611	
Mar 03, 2018 14:19	APPROVED	07A0A120-EC0D-4007-B19E-C50D79AA3611	Approved with three adjustments
Mar 06, 2018 18:47	APPROVED	07A0A120-EC0D-4007-B19E-C50D79AA3611	
Mar 06, 2018 18:53	APPROVED	07A0A120-EC0D-4007-B19E-C50D79AA3611	
Mar 06, 2018 19:03	APPROVED	07A0A120-EC0D-4007-B19E-C50D79AA3611	

Figure A.20: An example of a users history of reported lineups corresponding to a trust value of 0.71.

Reported At	Decision	Reporter Forza ID	Quality
Sep 30, 2017 14:20	APPROVED	872D8888-608A-469E-BFB3-F07C541AED87	
Sep 30, 2017 15:34	APPROVED	872D8888-608A-469E-BFB3-F07C541AED87	
Sep 30, 2017 15:37	APPROVED	872D8888-608A-469E-BFB3-F07C541AED87	
Oct 09, 2017 17:43	APPROVED	872D8888-608A-469E-BFB3-F07C541AED87	
Oct 18, 2017 17:35	DISCARDED	872D8888-608A-469E-BFB3-F07C541AED87	
Oct 18, 2017 17:38	DISCARDED	872D8888-608A-469E-BFB3-F07C541AED87	
Nov 05, 2017 15:38	APPROVED	872D8888-608A-469E-BFB3-F07C541AED87	
Nov 14, 2017 19:01	APPROVED	872D8888-608A-469E-BFB3-F07C541AED87	
Nov 26, 2017 18:45	APPROVED	872D8888-608A-469E-BFB3-F07C541AED87	
Nov 26, 2017 18:46	APPROVED	872D8888-608A-469E-BFB3-F07C541AED87	
Nov 26, 2017 18:47	APPROVED	872D8888-608A-469E-BFB3-F07C541AED87	
Jan 02, 2018 19:02	APPROVED	872D8888-608A-469E-BFB3-F07C541AED87	Approved with three adjustments
Jan 02, 2018 19:05	APPROVED	872D8888-608A-469E-BFB3-F07C541AED87	
Jan 04, 2018 17:06	DISCARDED	872D8888-608A-469E-BFB3-F07C541AED87	
Jan 07, 2018 9:57	APPROVED	872D8888-608A-469E-BFB3-F07C541AED87	Approved with one adjustment
Jan 07, 2018 10:00	APPROVED	872D8888-608A-469E-BFB3-F07C541AED87	
Jan 07, 2018 18:51	APPROVED	872D8888-608A-469E-BFB3-F07C541AED87	
Jan 07, 2018 18:53	APPROVED	872D8888-608A-469E-BFB3-F07C541AED87	

Figure A.21: An example of a users history of reported lineups corresponding to a trust value of 0.86.

Reported At	Decision	Reporter Forza ID	Quality
Sep 26, 2017 21:53	APPROVED	22D00AE8-F918-4897-B05F-7AA337CEA15E	
Oct 14, 2017 13:07	APPROVED	22D00AE8-F918-4897-B05F-7AA337CEA15E	
Nov 22, 2017 19:40	APPROVED	22D00AE8-F918-4897-B05F-7AA337CEA15E	
Dec 01, 2017 18:57	APPROVED	22D00AE8-F918-4897-B05F-7AA337CEA15E	
Dec 09, 2017 14:21	APPROVED	22D00AE8-F918-4897-B05F-7AA337CEA15E	
Dec 16, 2017 14:02	APPROVED	22D00AE8-F918-4897-B05F-7AA337CEA15E	
Jan 07, 2018 11:26	APPROVED	22D00AE8-F918-4897-B05F-7AA337CEA15E	
Jan 13, 2018 14:11	APPROVED	22D00AE8-F918-4897-B05F-7AA337CEA15E	
Jan 20, 2018 14:17	APPROVED	22D00AE8-F918-4897-B05F-7AA337CEA15E	Approved with three adjustments
Jan 30, 2018 18:51	DISCARDED	22D00AE8-F918-4897-B05F-7AA337CEA15E	
Feb 18, 2018 15:36	APPROVED	22D00AE8-F918-4897-B05F-7AA337CEA15E	
Feb 24, 2018 14:08	APPROVED	22D00AE8-F918-4897-B05F-7AA337CEA15E	

Figure A.22: An example of a users history of reported lineups corresponding to a trust value of 0.94.

B

Material used in the validation sessions based on users trust values

The following material was used in the test session that took place in the **first** iteration. Note that the examples of users trust values, is shown in an increased order, and not in the order in which they were shown to the participants in the test sessions.

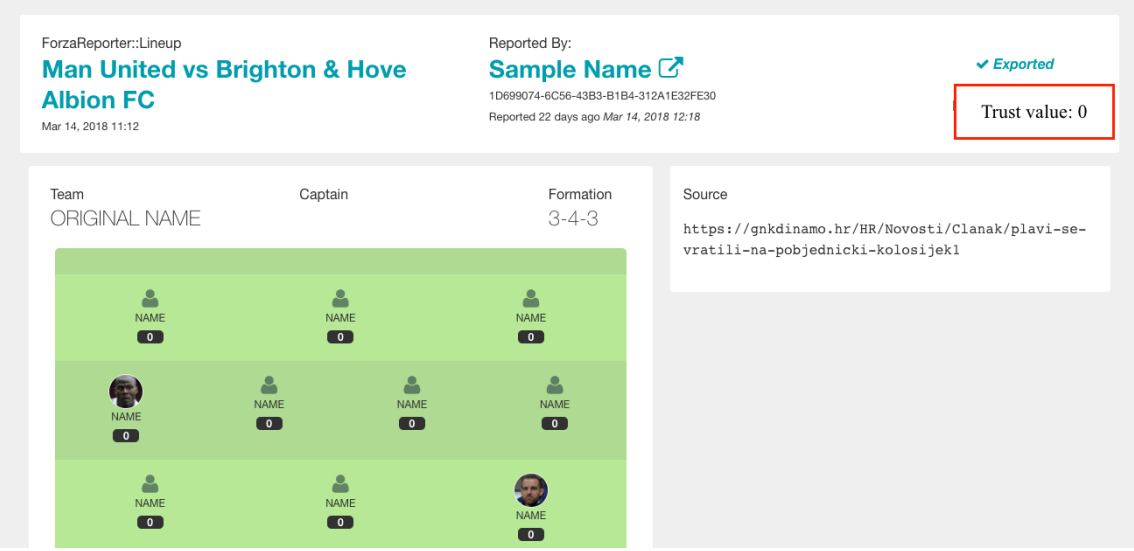


Figure B.1: An example of a user with trust value 0, displayed in admin.

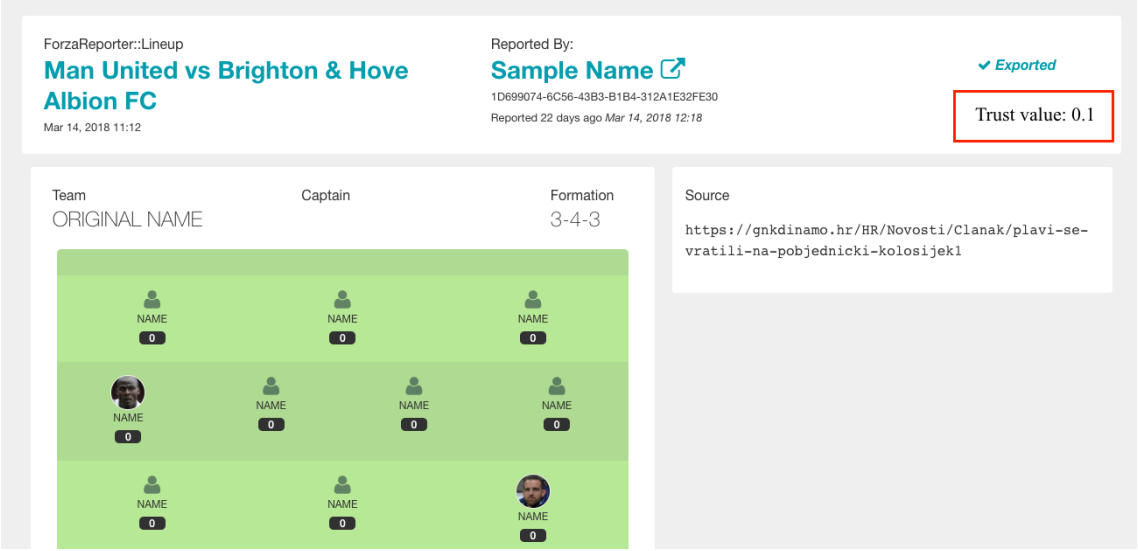


Figure B.2: An example of a user with trust value 0.1, displayed in admin.

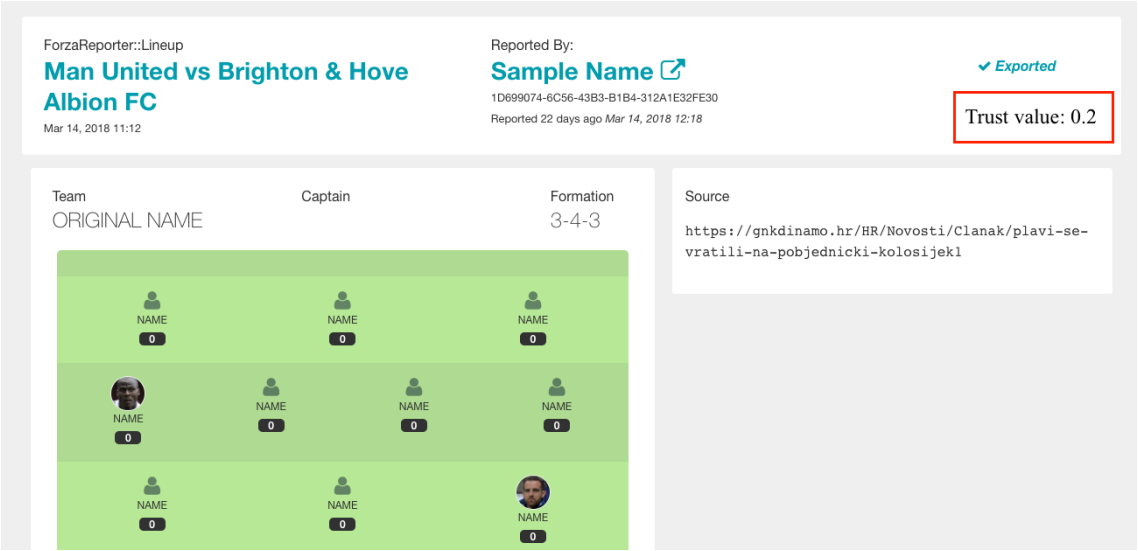


Figure B.3: An example of a user with trust value 0.2, displayed in admin.

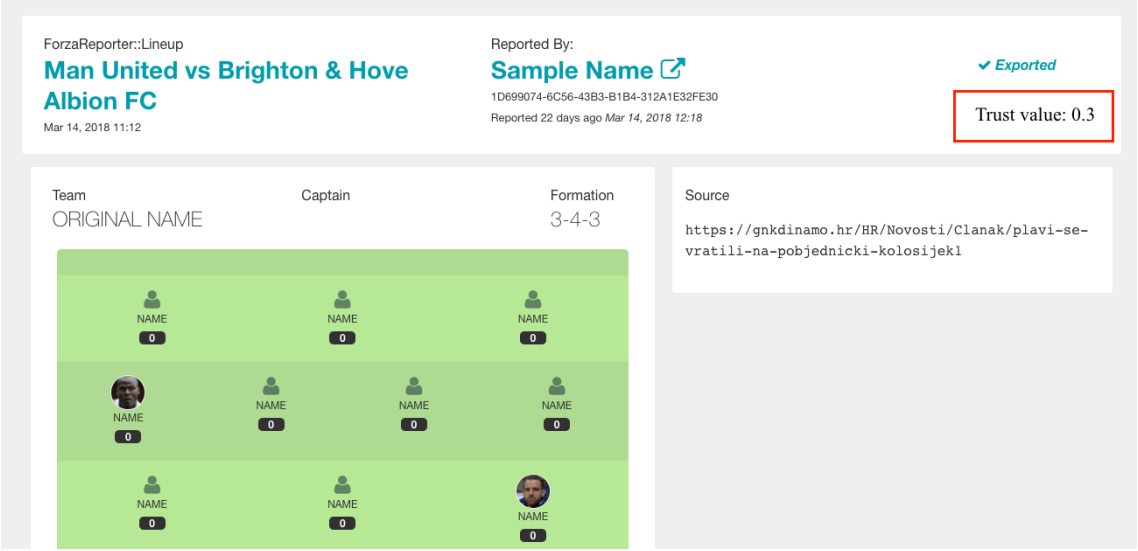


Figure B.4: An example of a user with trust value 0.3, displayed in admin.

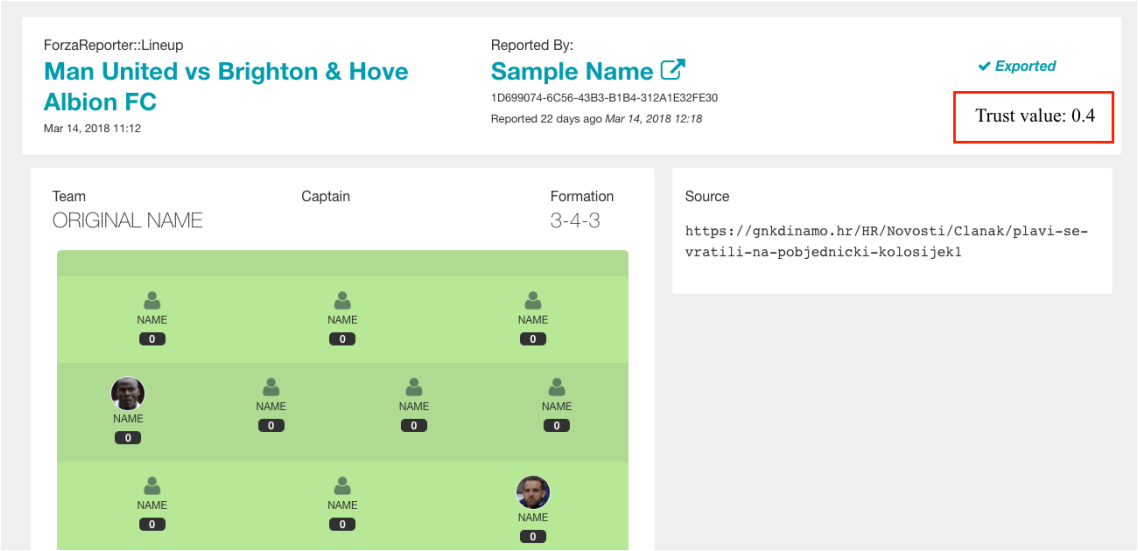


Figure B.5: An example of a user with trust value 0.4, displayed in admin.

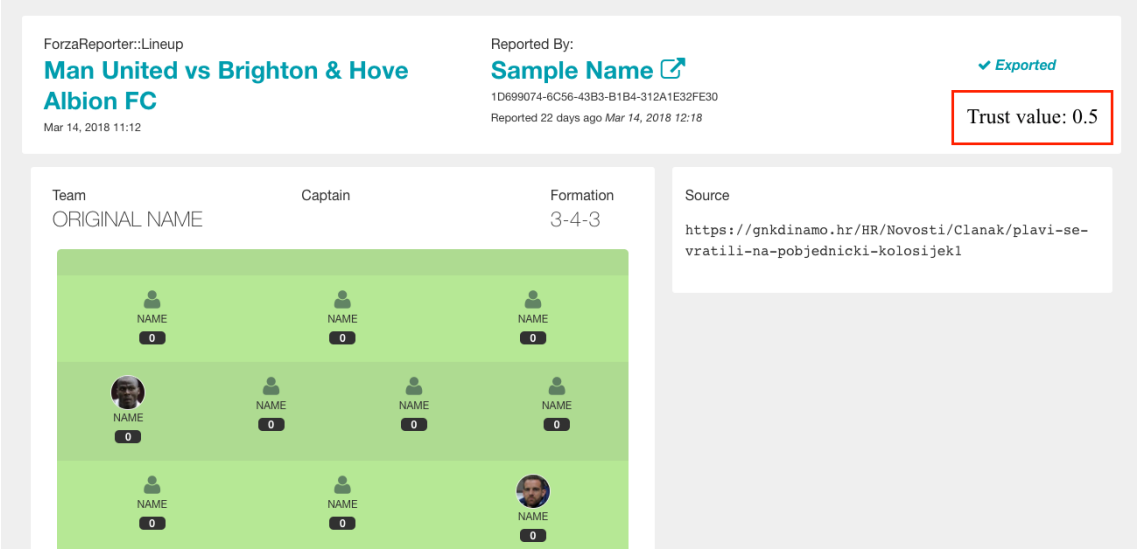


Figure B.6: An example of a user with trust value 0.5, displayed in admin.

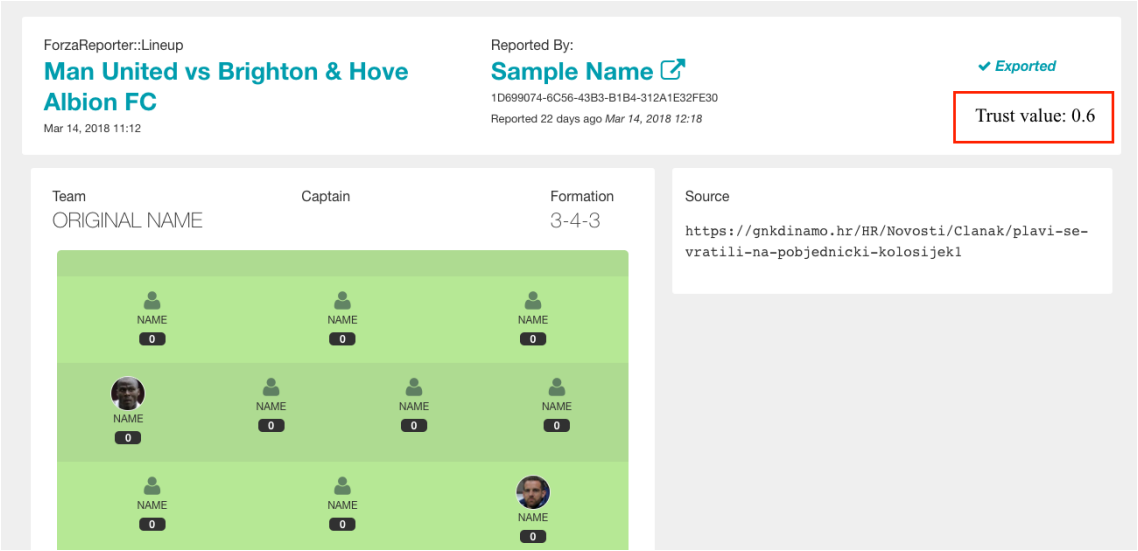


Figure B.7: An example of a user with trust value 0.6, displayed in admin.

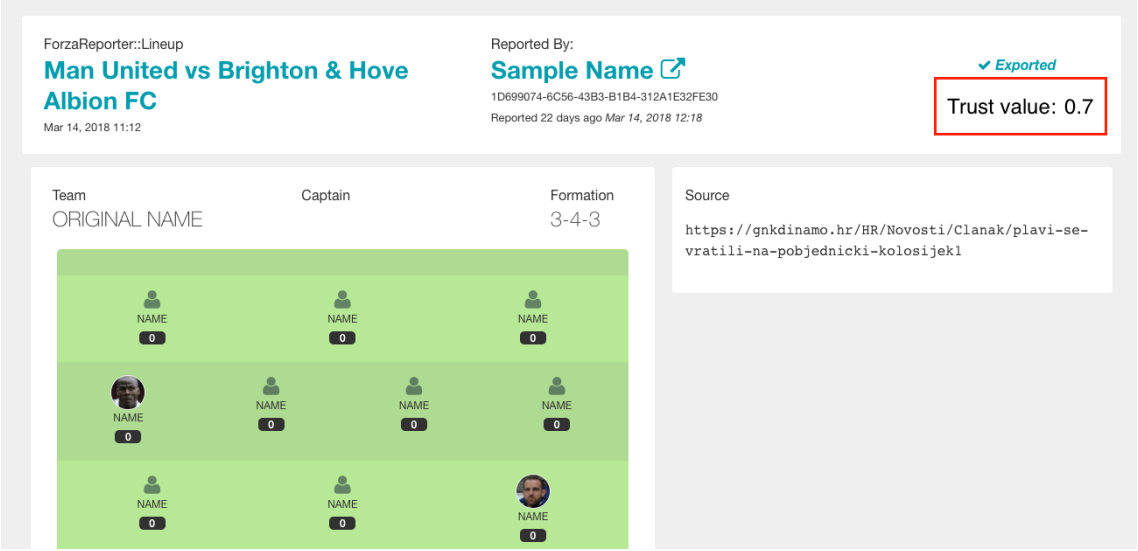


Figure B.8: An example of a user with trust value 0.7, displayed in admin.

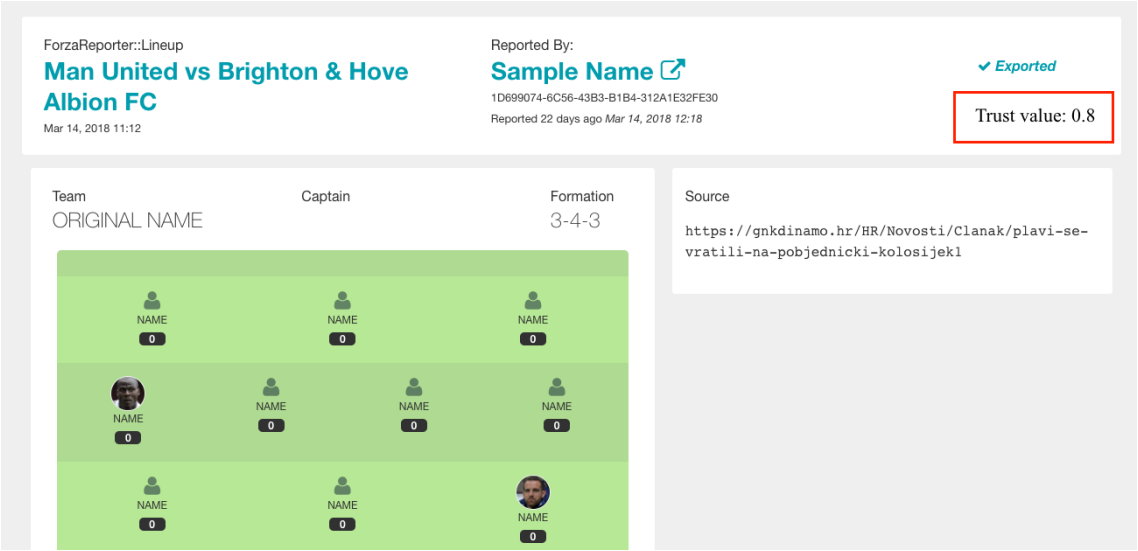


Figure B.9: An example of a user with trust value 0.8, displayed in admin.

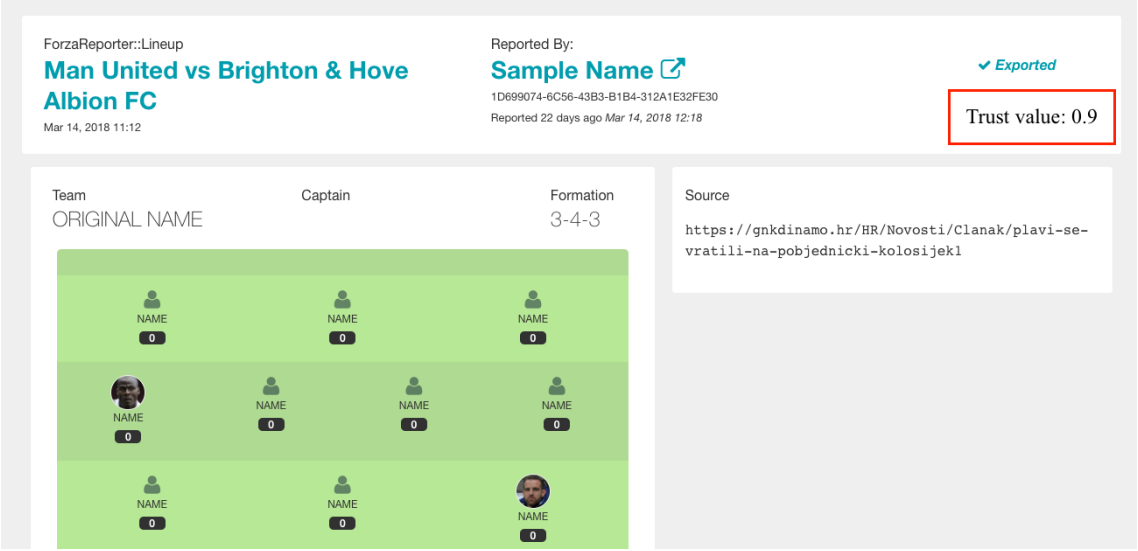


Figure B.10: An example of a user with trust value 0.9, displayed in admin.

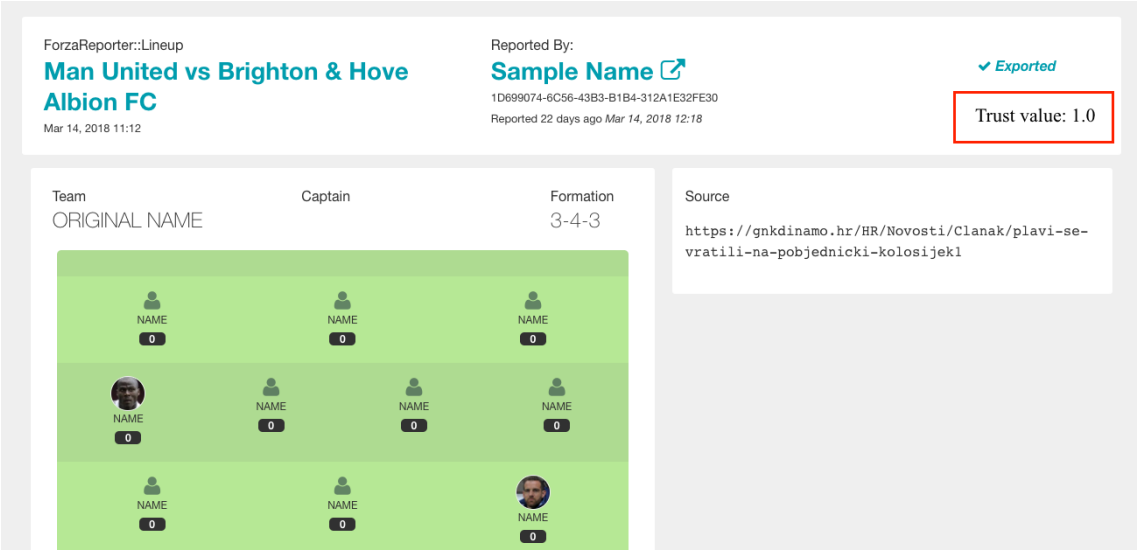


Figure B.11: An example of a user with trust value 1.0, displayed in admin.

The following material was used in the test session that took place in the **second** iteration. Note that the examples of users trust values, is shown in an increased order, and not in the order in which they were shown to the participants in the test sessions.

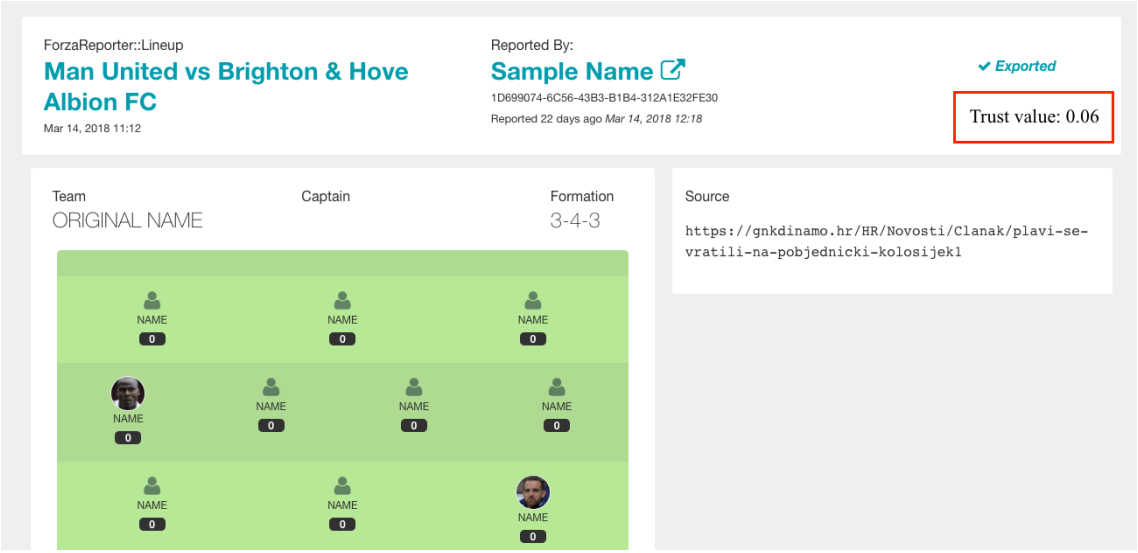


Figure B.12: An example of a user with trust value 0.06, displayed in admin.

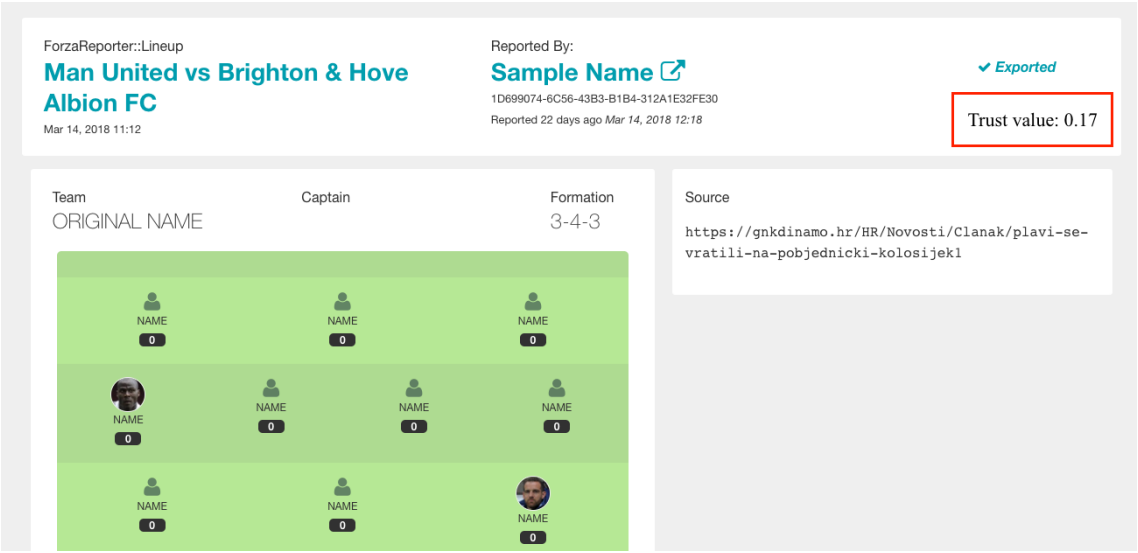


Figure B.13: An example of a user with trust value 0.17, displayed in admin.

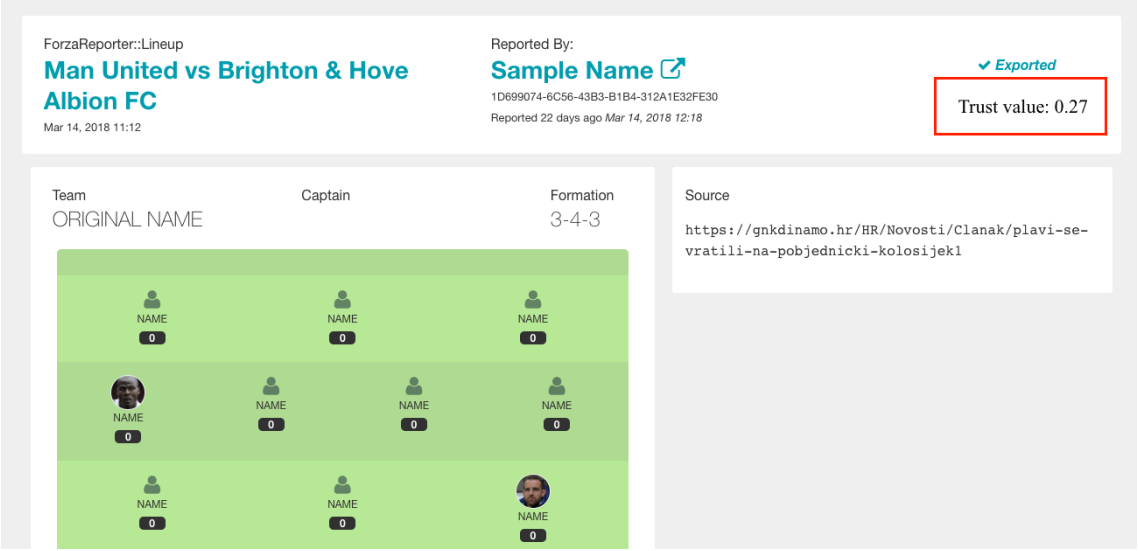


Figure B.14: An example of a user with trust value 0.27, displayed in admin.

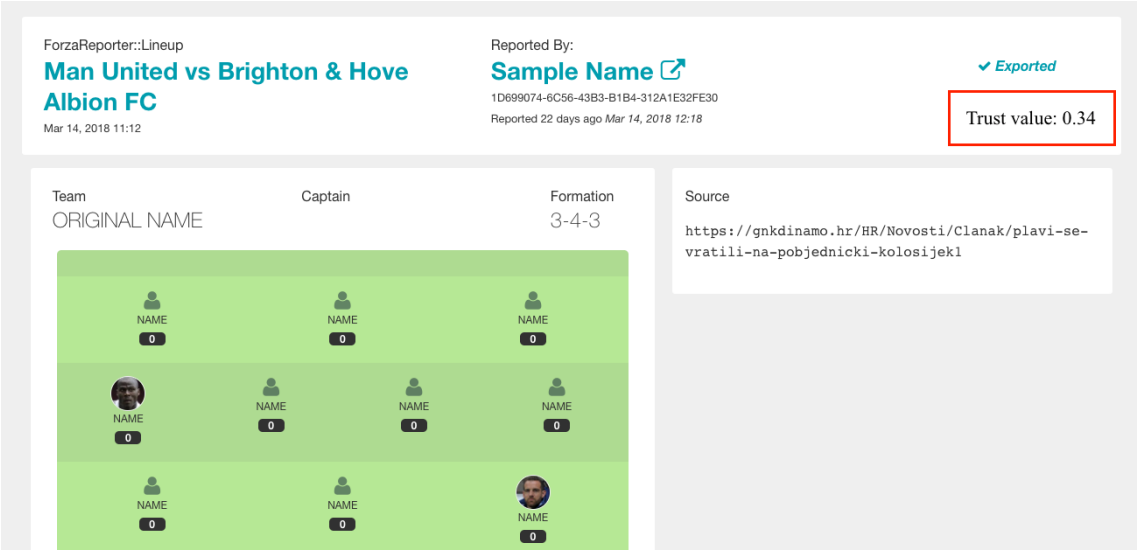


Figure B.15: An example of a user with trust value 0.34, displayed in admin.

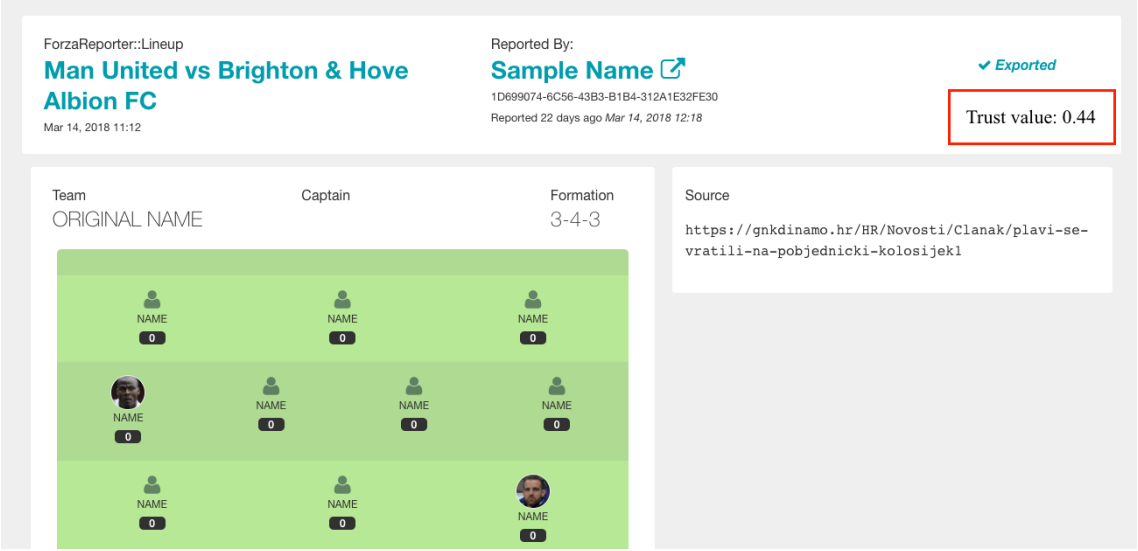


Figure B.16: An example of a user with trust value 0.44, displayed in admin.

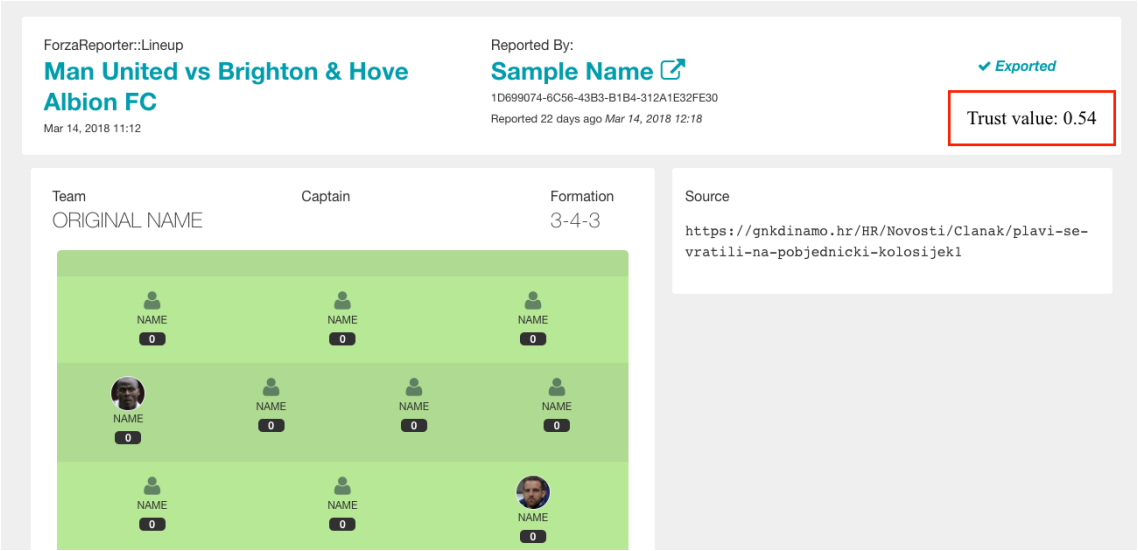


Figure B.17: An example of a user with trust value 0.54, displayed in admin.

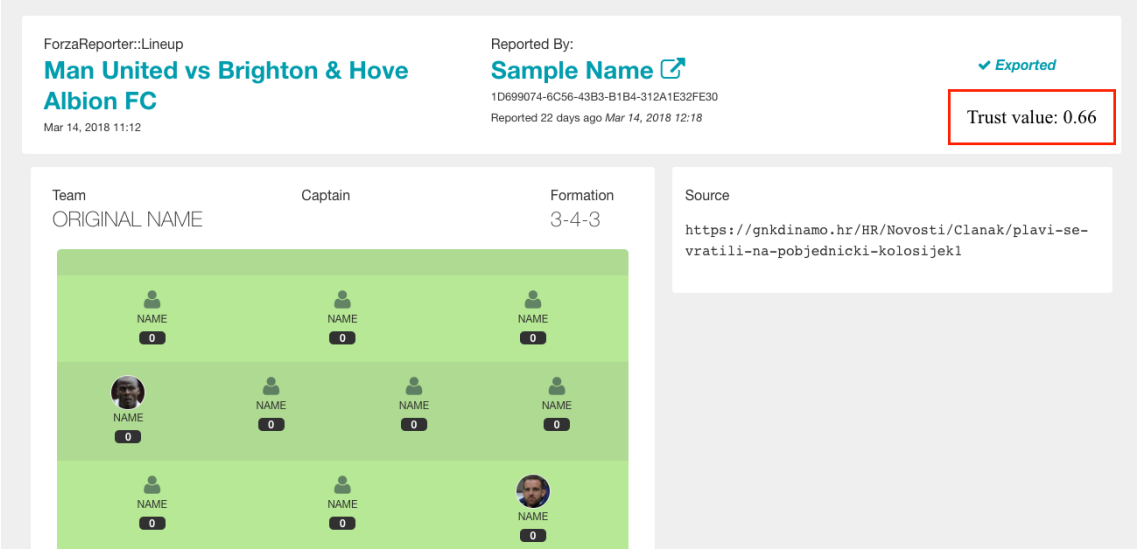


Figure B.18: An example of a user with trust value 0.66, displayed in admin.

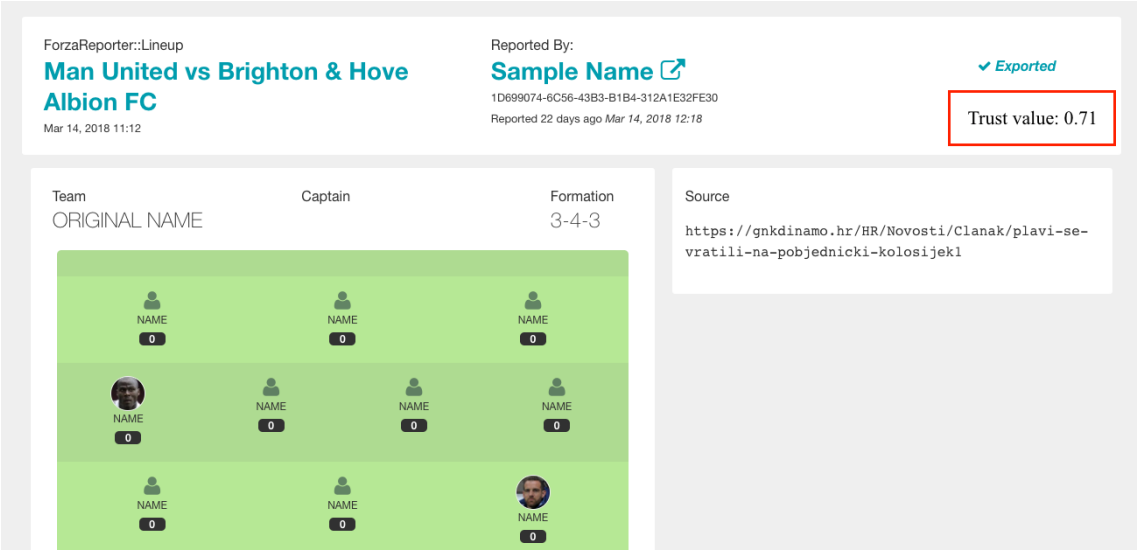


Figure B.19: An example of a user with trust value 0.71, displayed in admin.

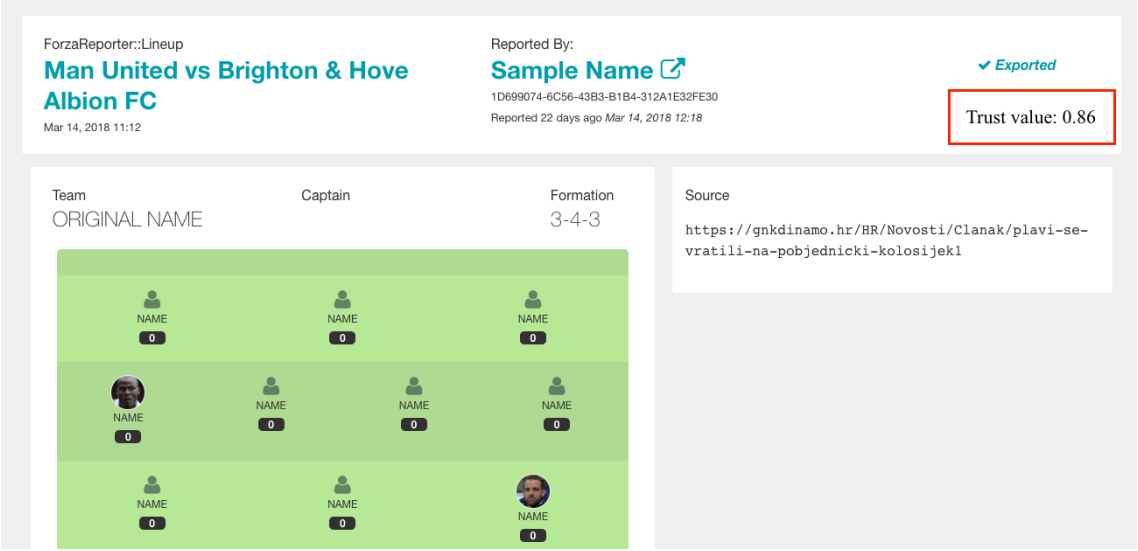


Figure B.20: An example of a user with trust value 0.86, displayed in admin.

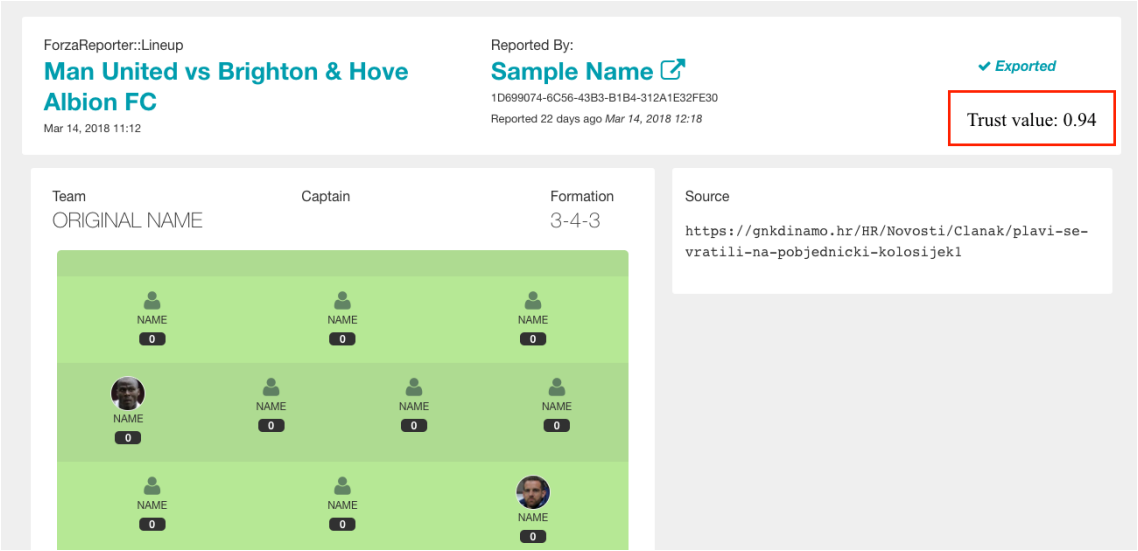


Figure B.21: An example of a user with trust value 0.94, displayed in admin.

C

Material used in the validation sessions for evaluating the trust value together with other metrics

In this appendix, all material used in the third iteration is presented. The material used in the first three validation sessions, where the trust value was evaluated together with users approval rate, users approval rate together with the number of reported lineups, and the number of approved lineups out of the five last reported ones, is presented in their own sections, in the given order. The material used in the final validation session, where the approval rate was evaluated together with number of reported lineups, is presented in the last section.

C.1 Trust value together with users' approval rate

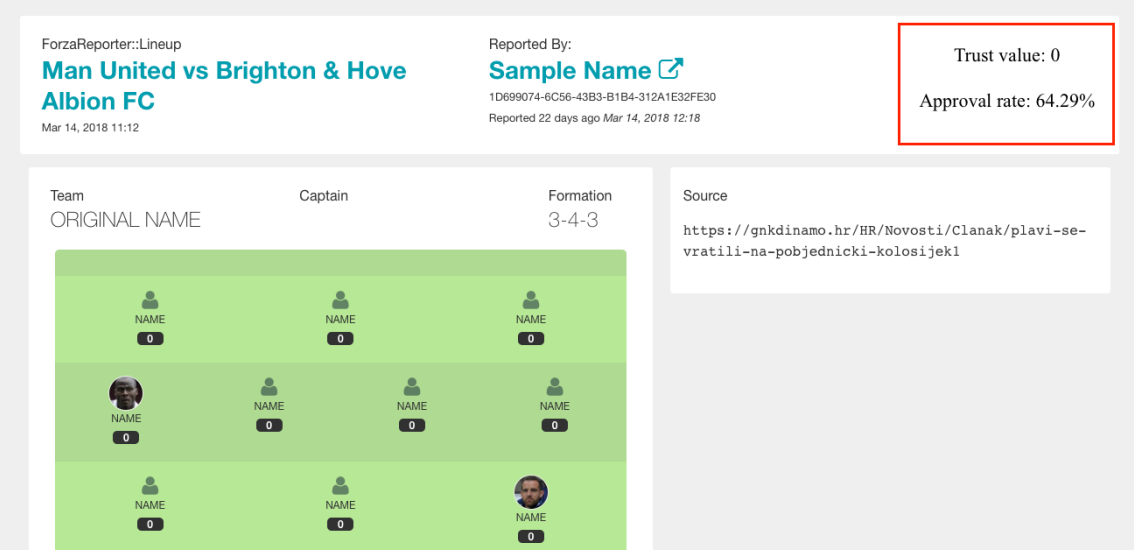


Figure C.1: An example of a user with trust value: 0, approval rate: 64.29%

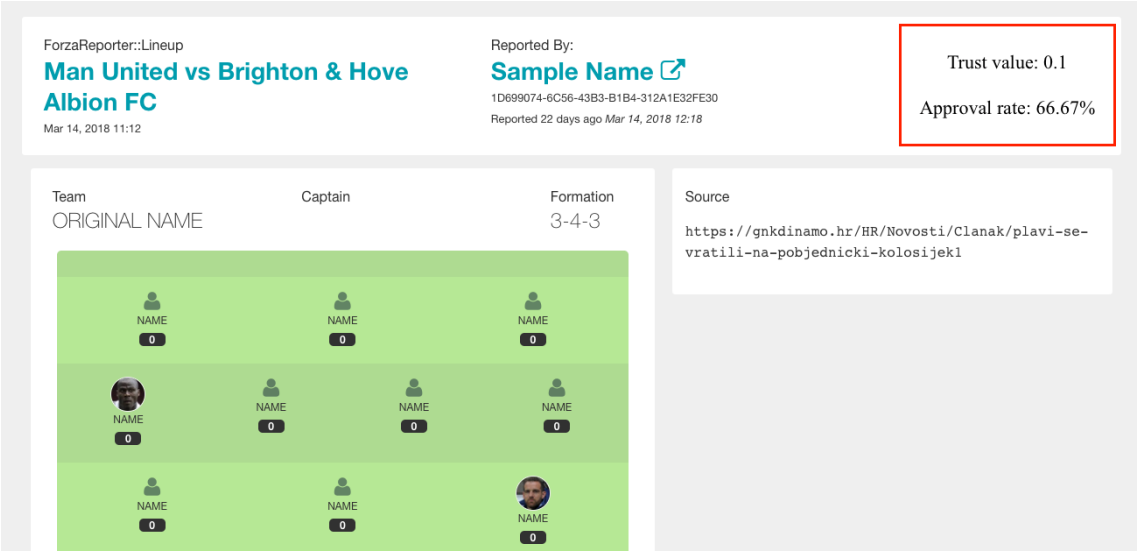


Figure C.2: An example of a user with trust value: 0.1, approval rate: 66.67%

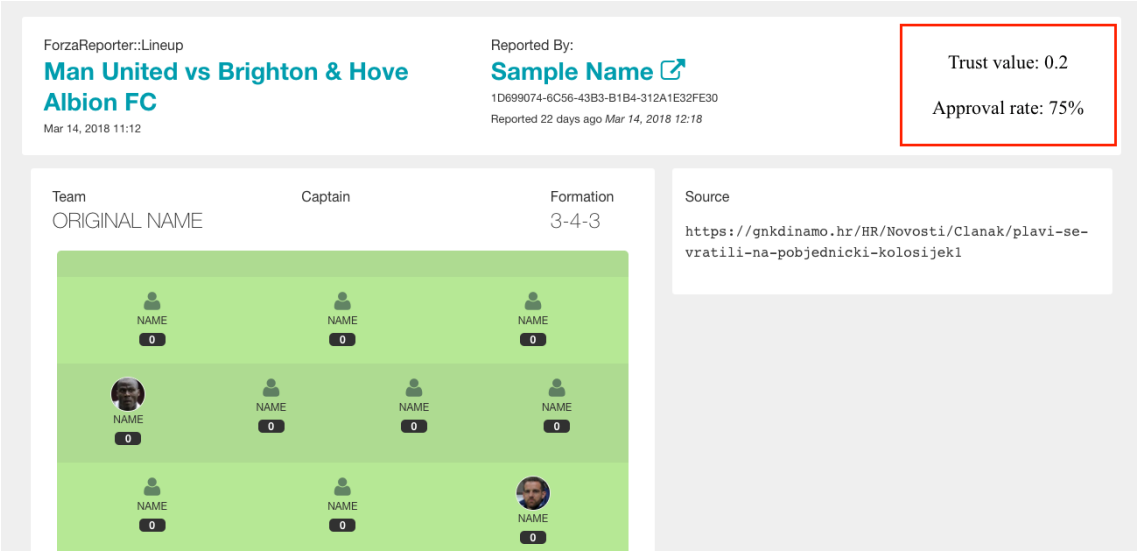


Figure C.3: An example of a user with trust value: 0.2, approval rate: 75%

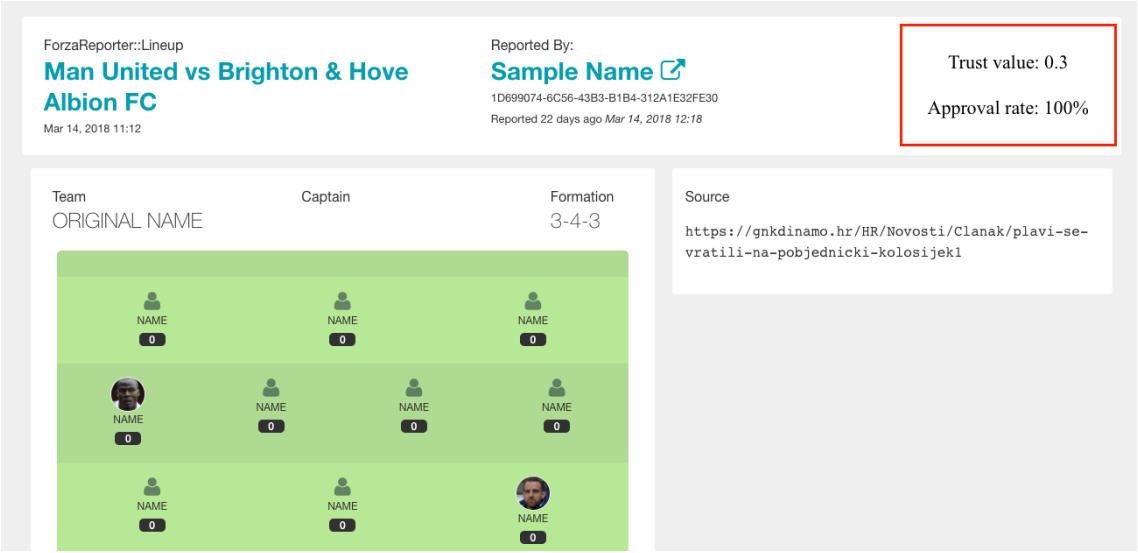


Figure C.4: An example of a user with trust value: 0.3, approval rate: 100%

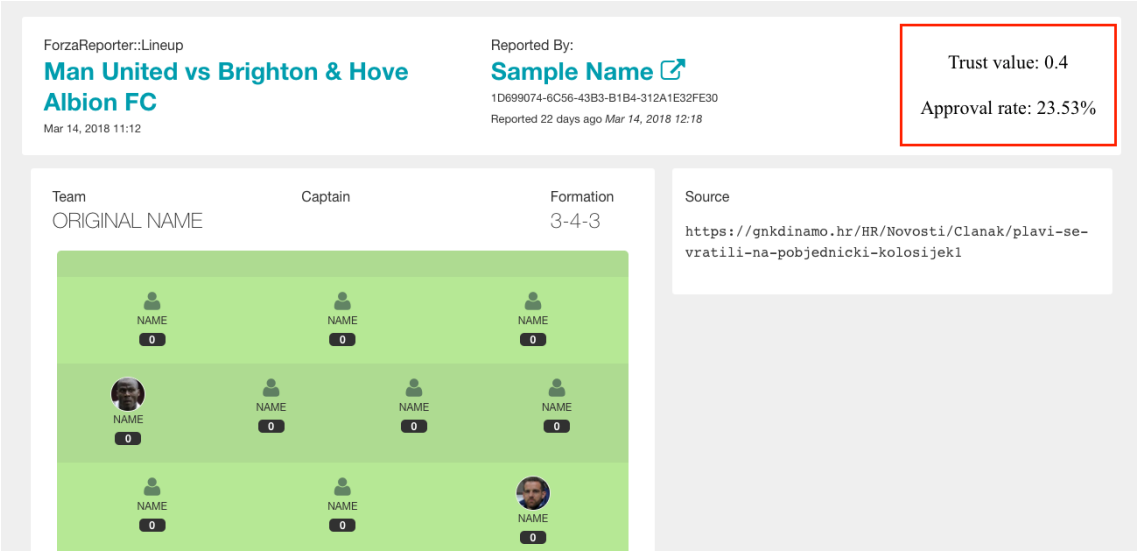


Figure C.5: An example of a user with trust value: 0.4, approval rate: 23.53%

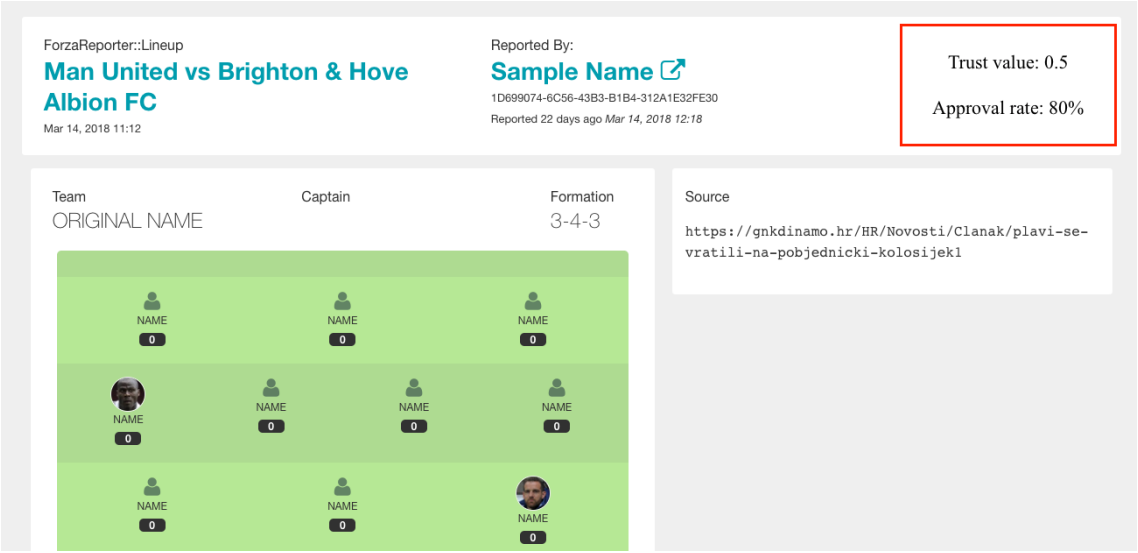


Figure C.6: An example of a user with trust value: 0.5, approval rate: 80%

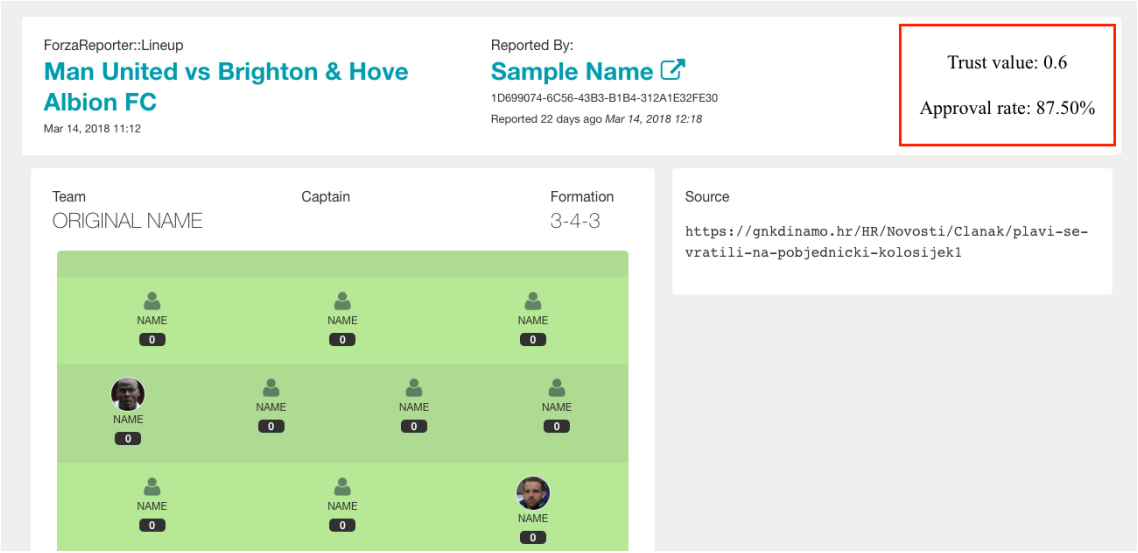


Figure C.7: An example of a user with trust value: 0.6, approval rate: 87.50%

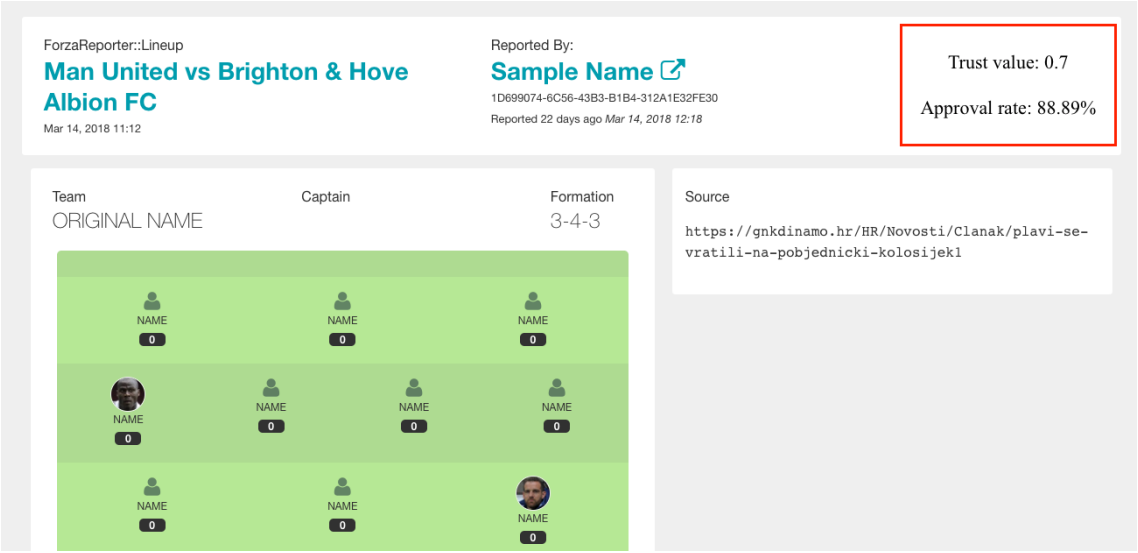


Figure C.8: An example of a user with trust value: 0.7, approval rate: 88.89%

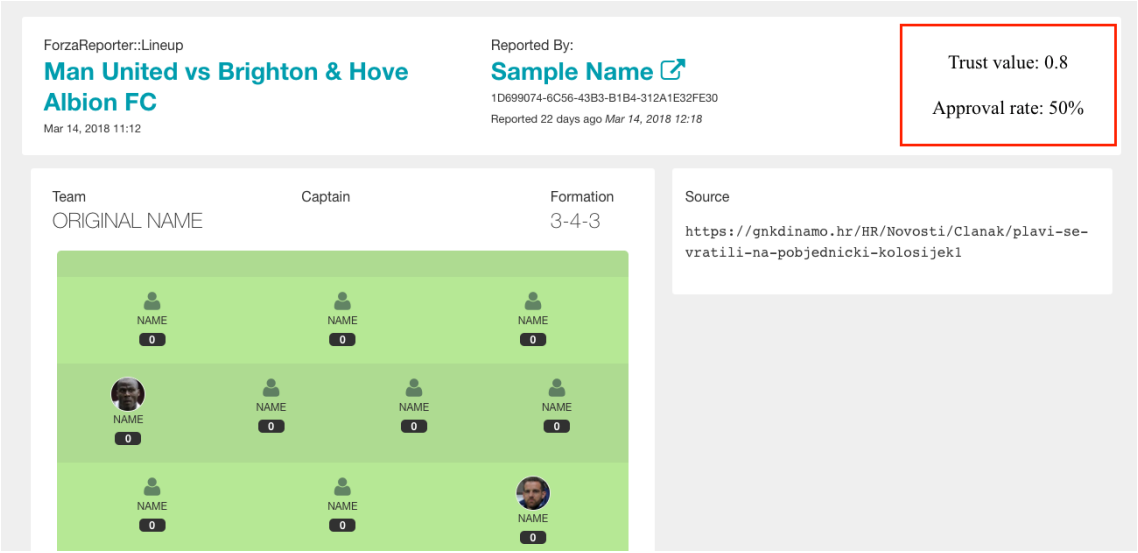


Figure C.9: An example of a user with trust value: 0.8, approval rate: 50%

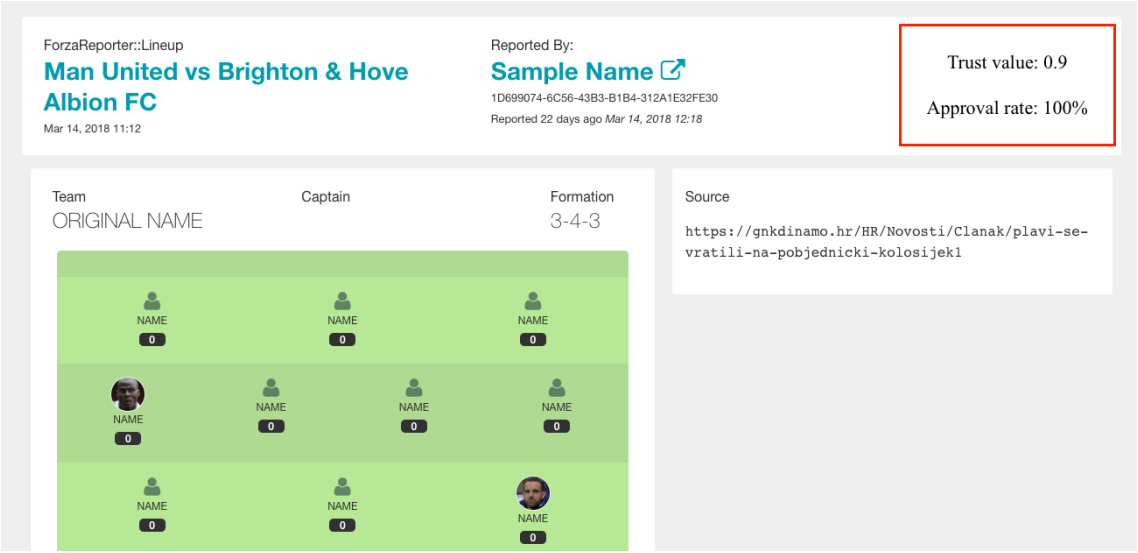


Figure C.10: An example of a user with trust value: 0.9, approval rate: 100%

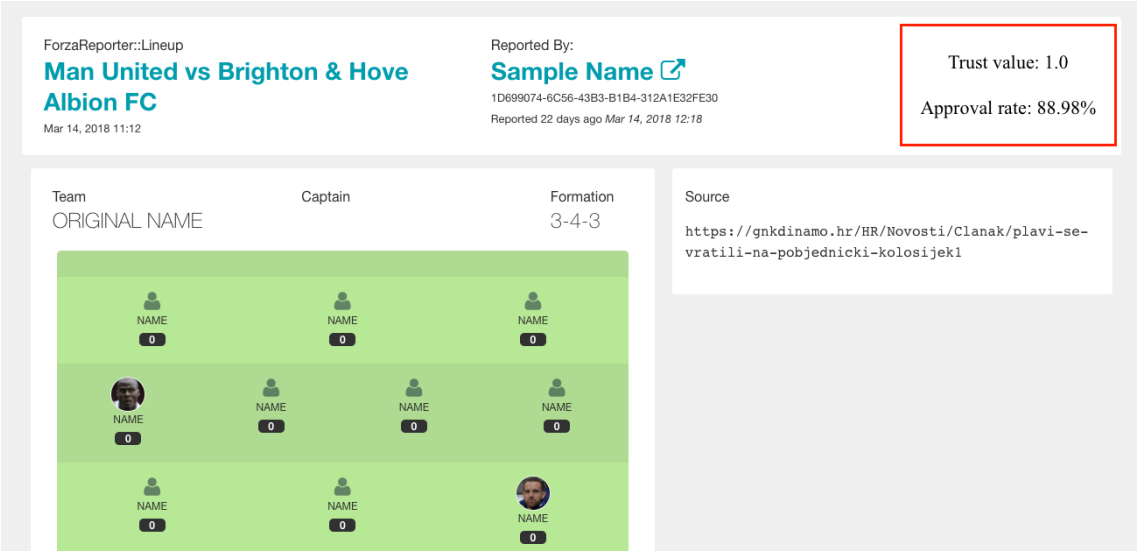


Figure C.11: An example of a user with trust value: 1.0, approval rate: 88.98%

C.2 Trust value together with users' approval rate and number of reported lineups.

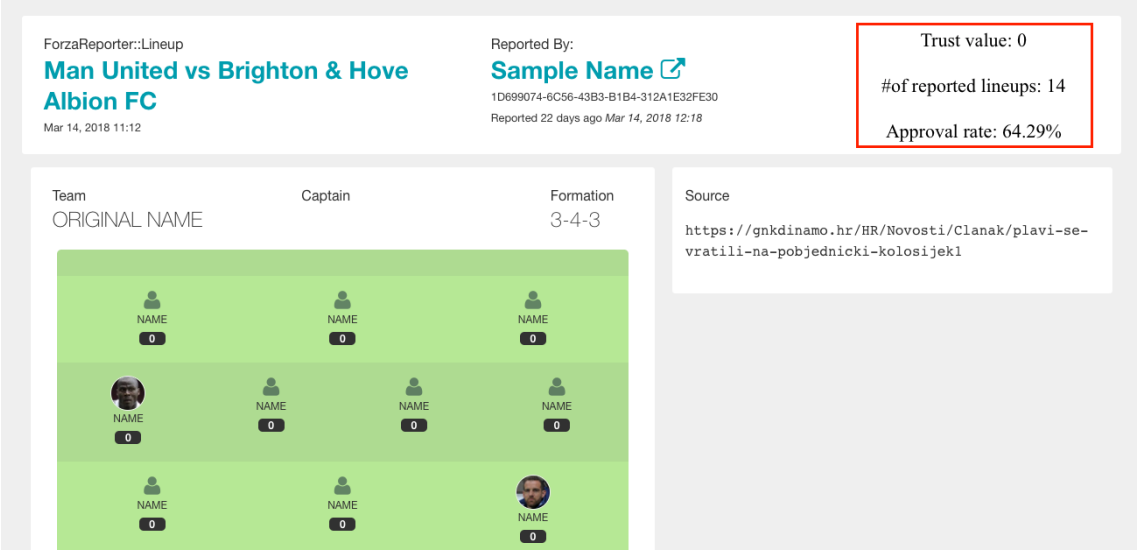


Figure C.12: An example of a user with trust value: 0, approval rate: 64.29%, #reported lineups: 14.

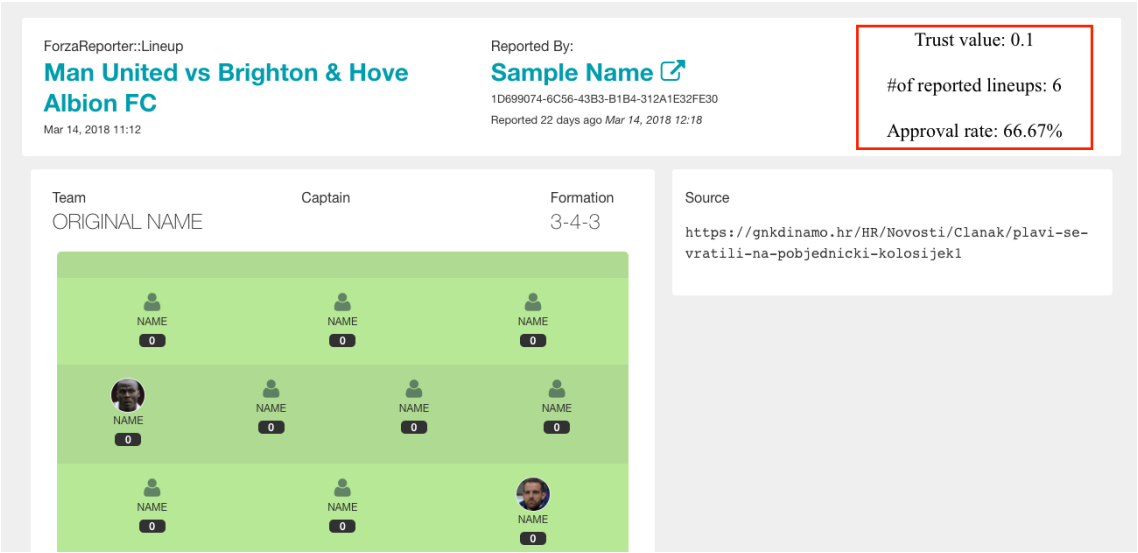


Figure C.13: An example of a user with trust value: 0.1, approval rate: 66.67%, #reported lineups: 6.

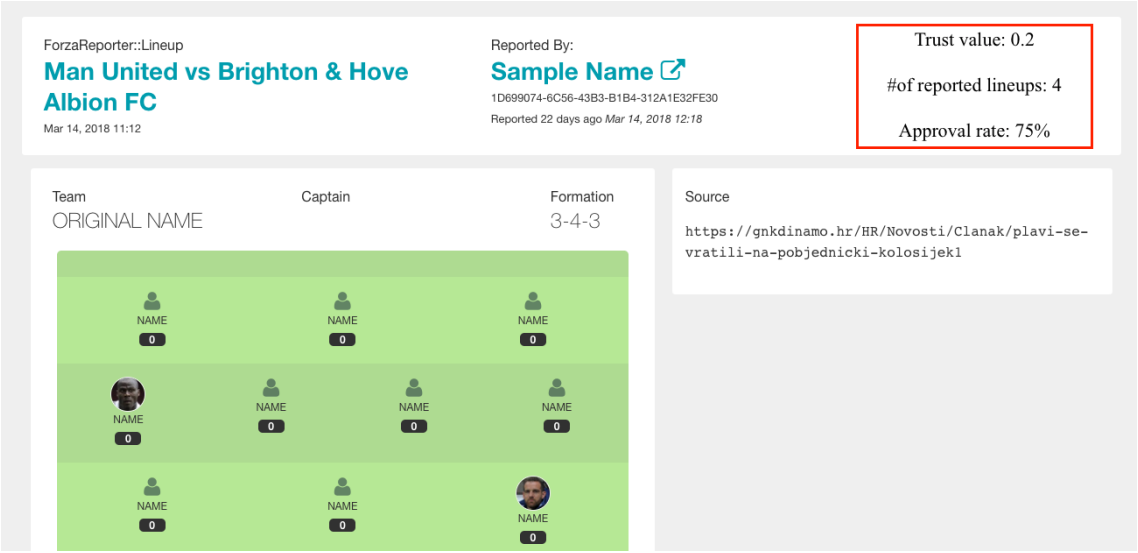


Figure C.14: An example of a user with trust value: 0.2, approval rate: 75%, #reported lineups: 4.

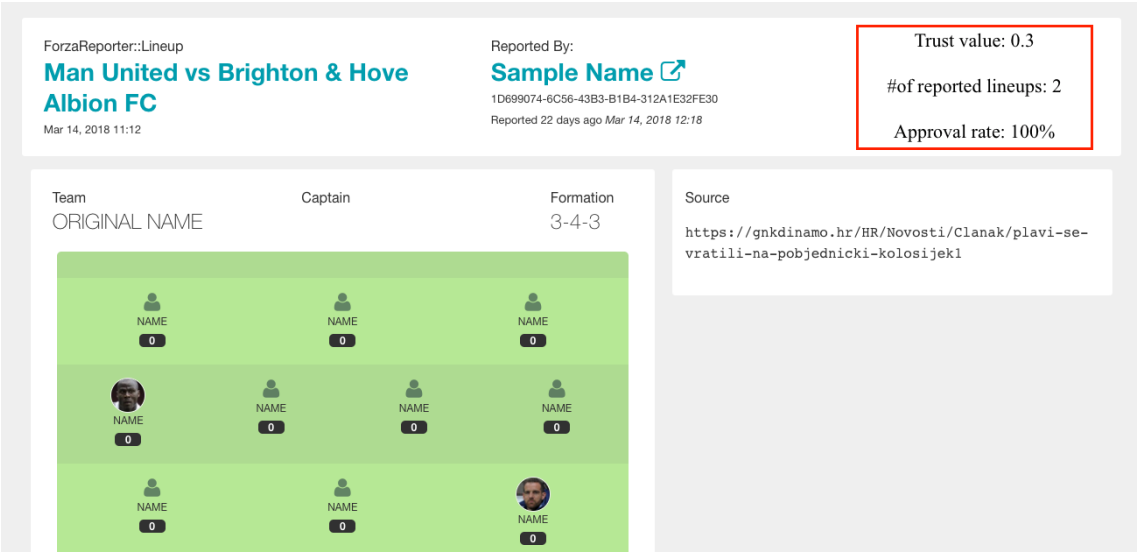


Figure C.15: An example of a user with trust value: 0.3, approval rate: 100%, #reported lineups: 2.

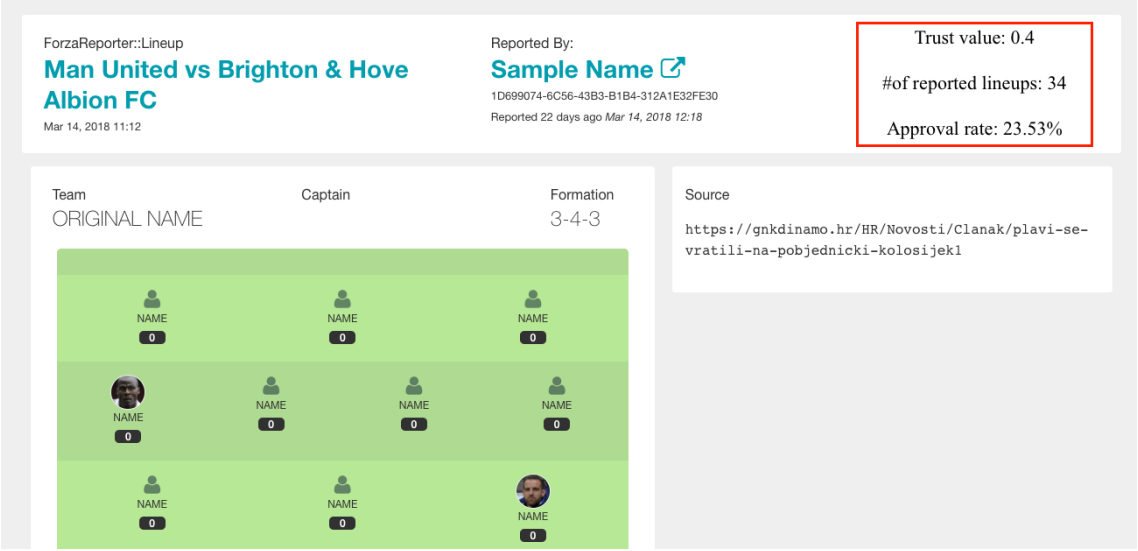


Figure C.16: An example of a user with trust value: 0.4, approval rate: 23.53%, #reported lineups: 34.

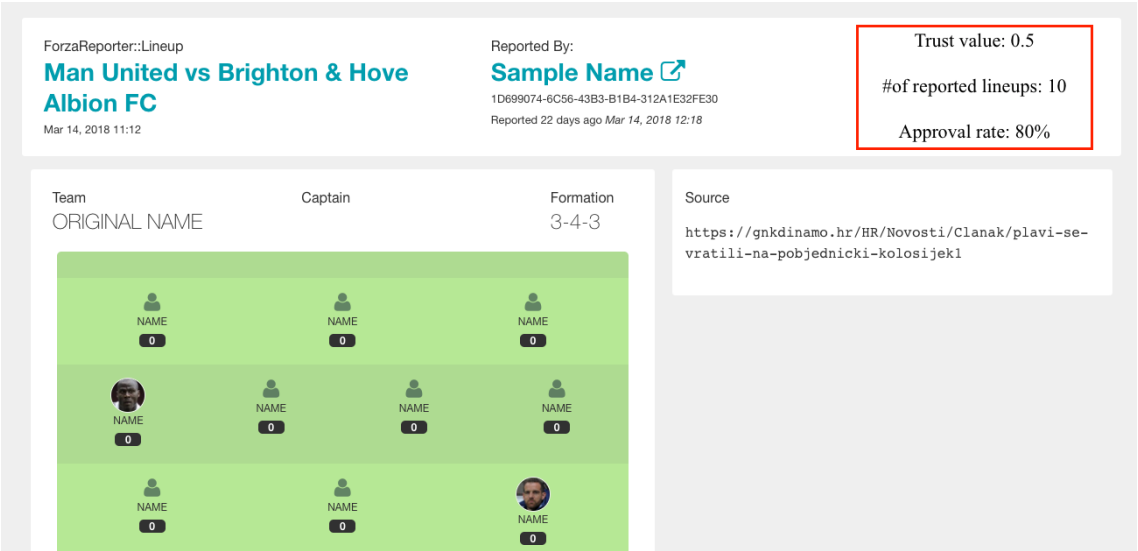


Figure C.17: An example of a user with trust value: 0.5, approval rate: 80%, #reported lineups: 10.

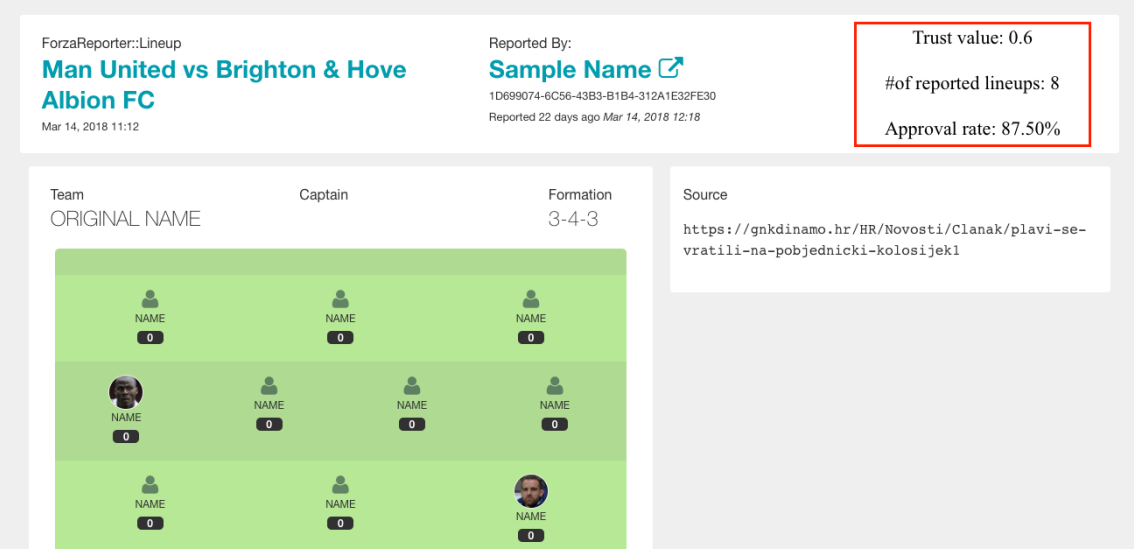


Figure C.18: An example of a user with trust value: 0.6, approval rate: 87.50%, #reported lineups: 8.

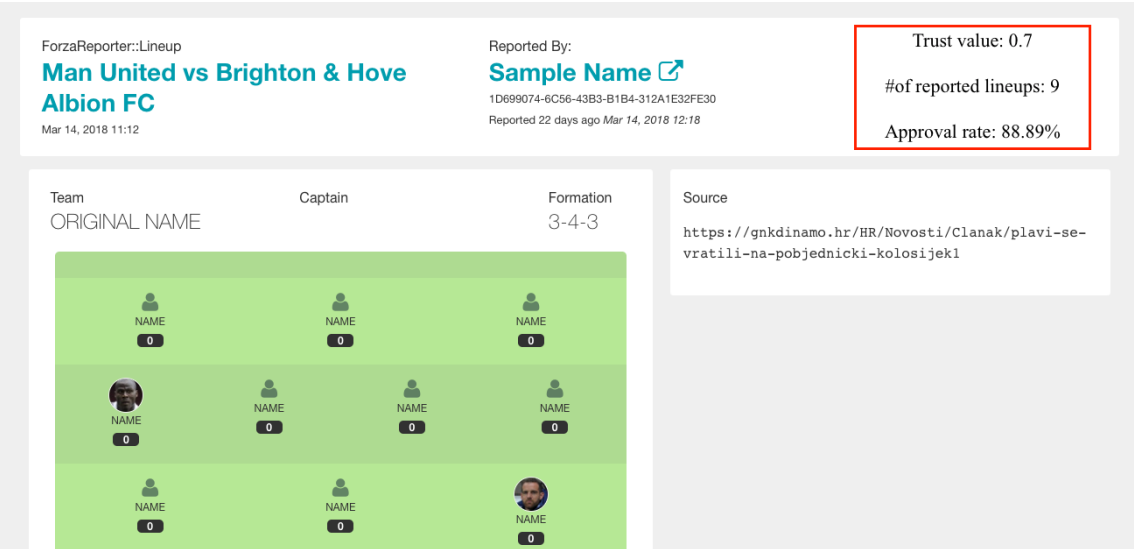


Figure C.19: An example of a user with trust value: 0.7, approval rate: 88.89%, #reported lineups: 9.

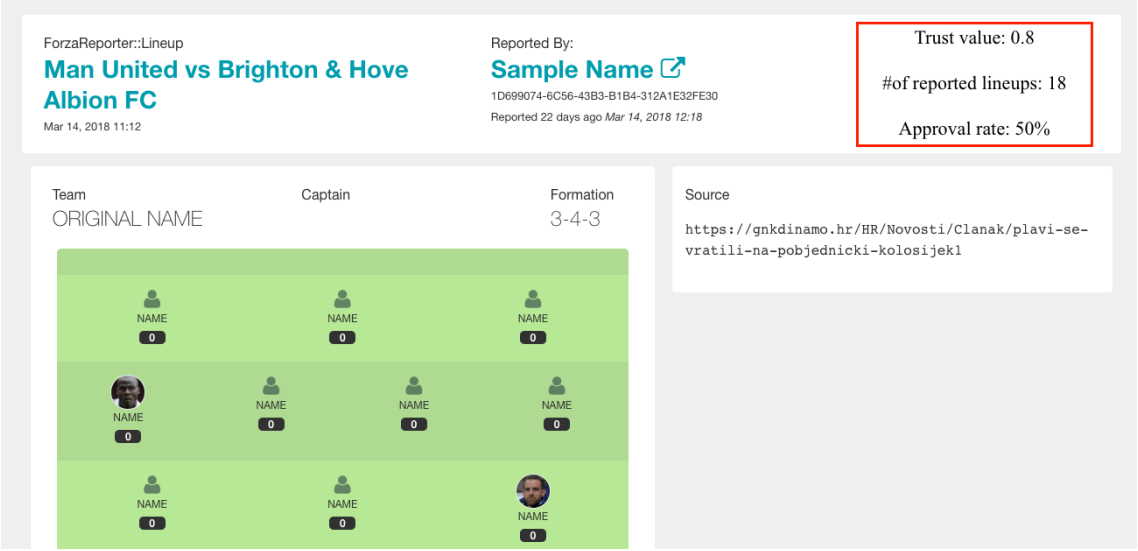


Figure C.20: An example of a user with trust value: 0.8, approval rate: 50%, #reported lineups: 18.

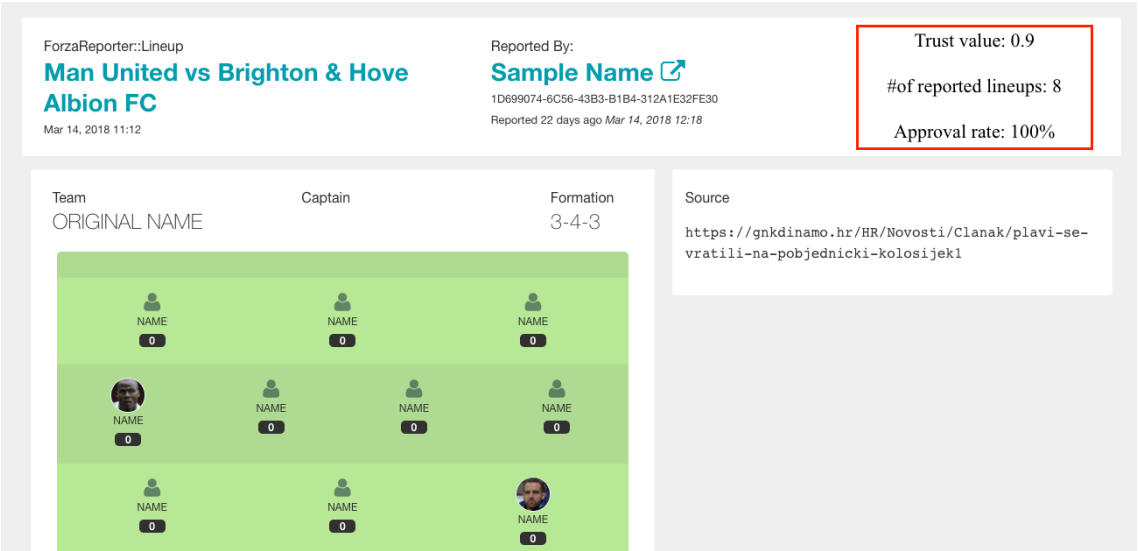


Figure C.21: An example of a user with trust value: 0.9, approval rate: 100%, #reported lineups: 8.

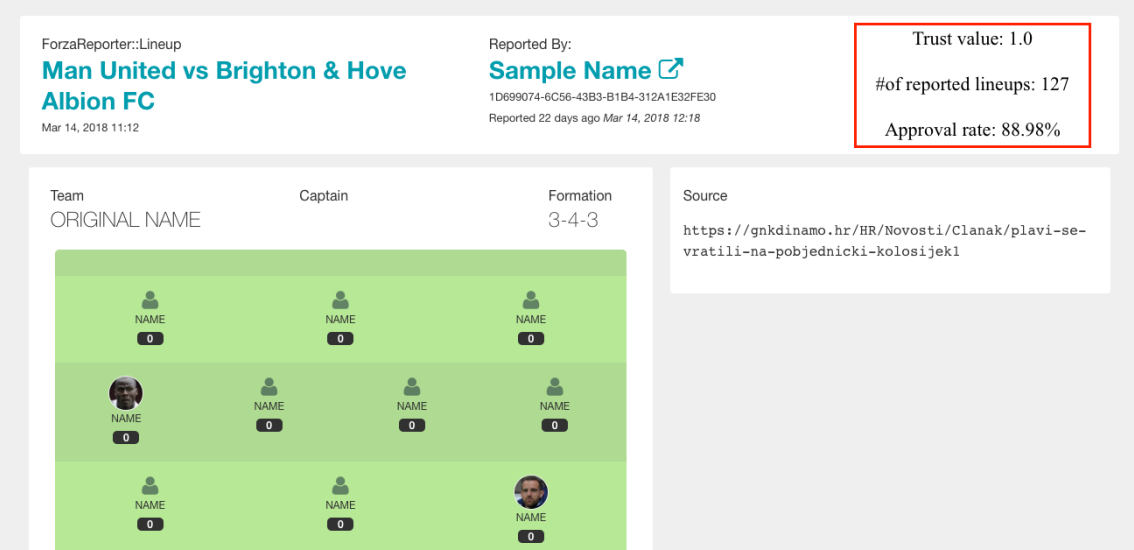


Figure C.22: An example of a user with trust value: 1.0, approval rate: 88.98%, #reported lineups: 127.

C.3 Trust value together with number of approved lineups out of the last five.

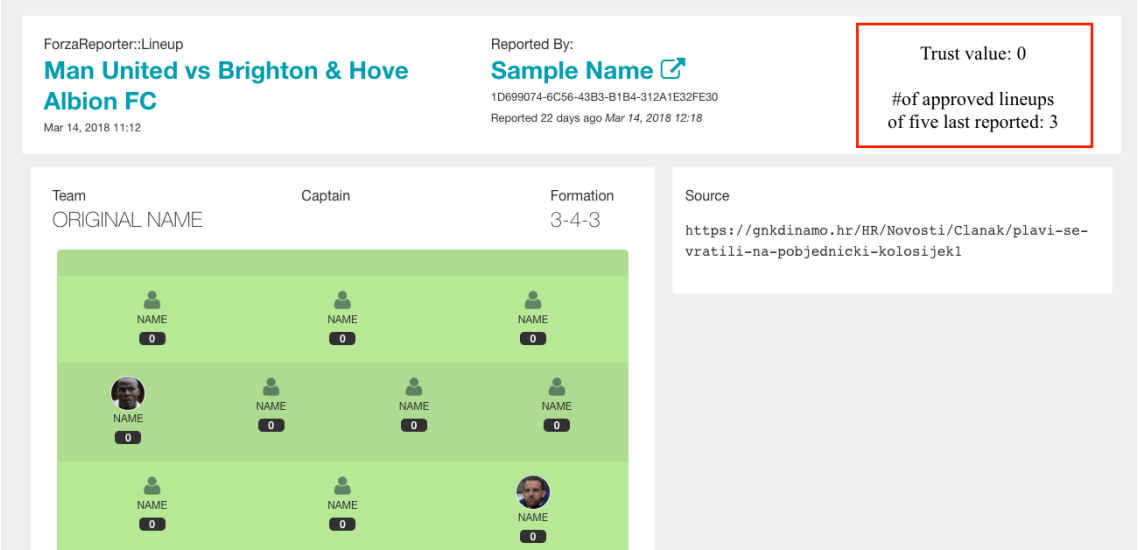


Figure C.23: An example of a user with trust value: 0, #approved lineups out of the last five: 3.

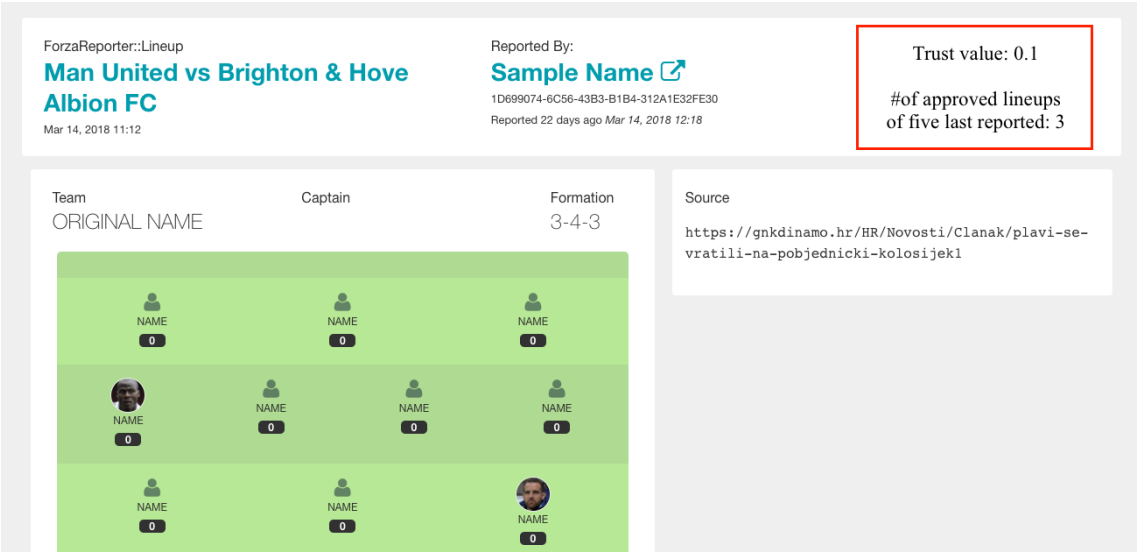


Figure C.24: An example of a user with trust value: 0.1, #approved lineups out of the last five: 3.

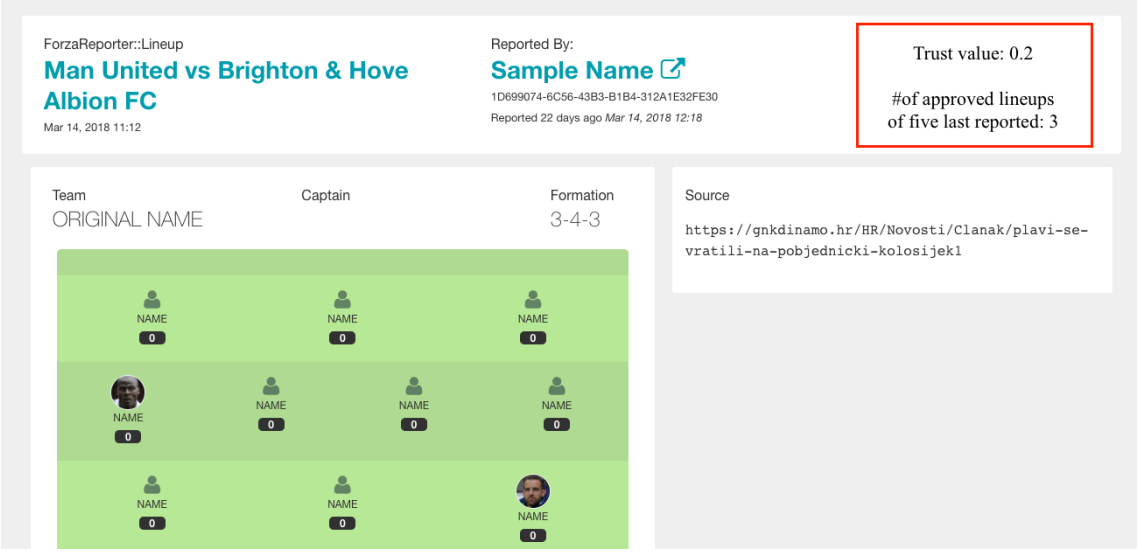


Figure C.25: An example of a user with trust value: 0.2, #approved lineups out of the last five: 3.

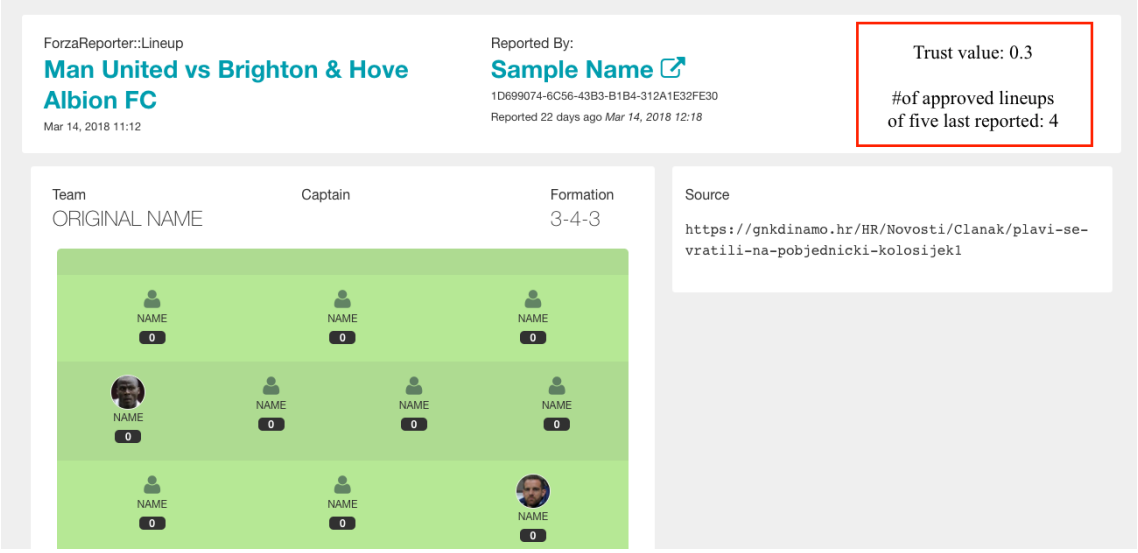


Figure C.26: An example of a user with trust value: 0.3, #approved lineups out of the last five: 4.

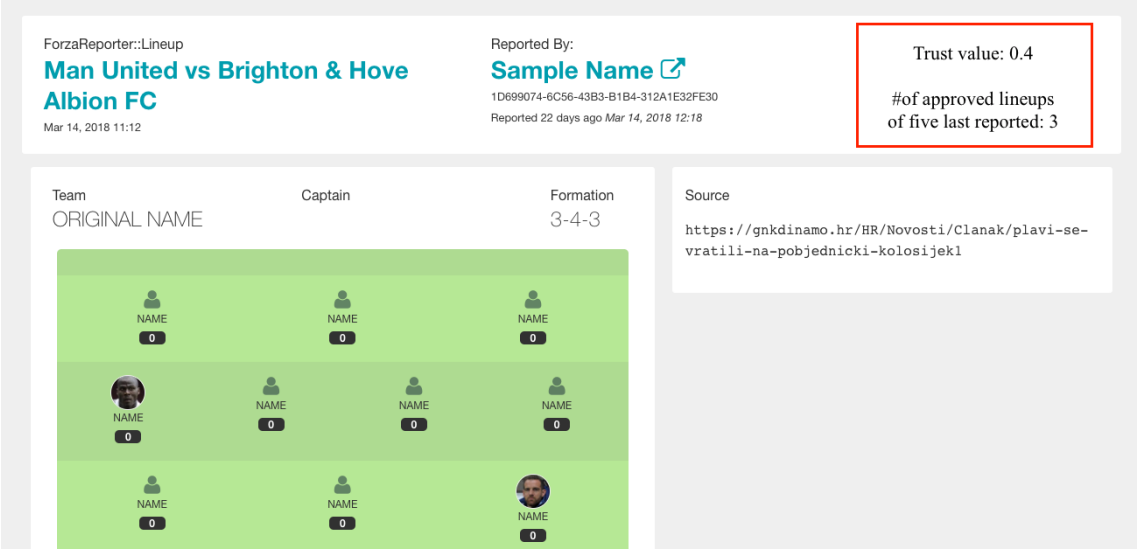


Figure C.27: An example of a user with trust value: 0.4, #approved lineups out of the last five: 3.

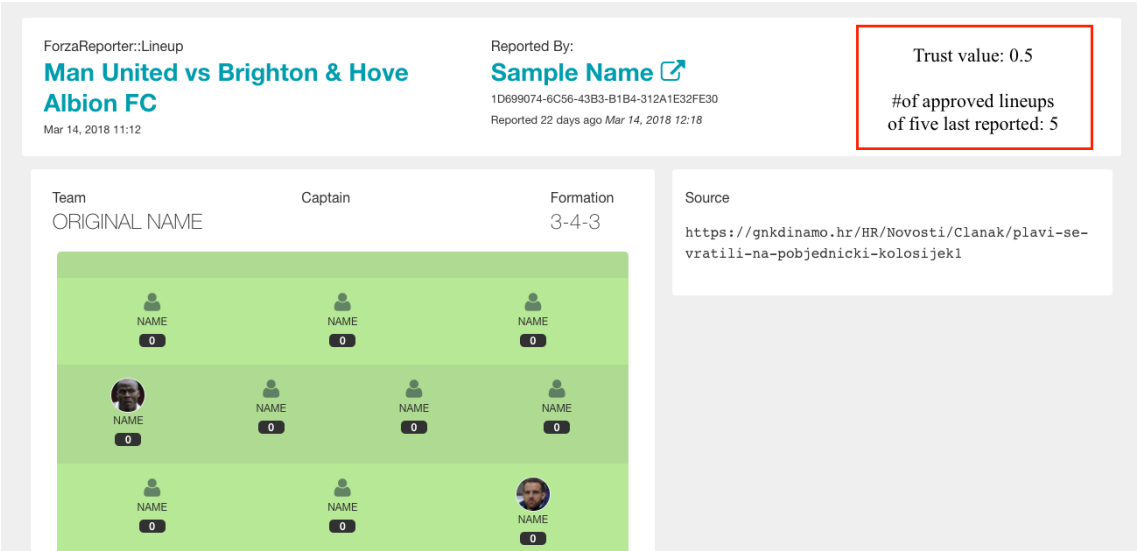


Figure C.28: An example of a user with trust value: 0.5, #approved lineups out of the last five: 5.

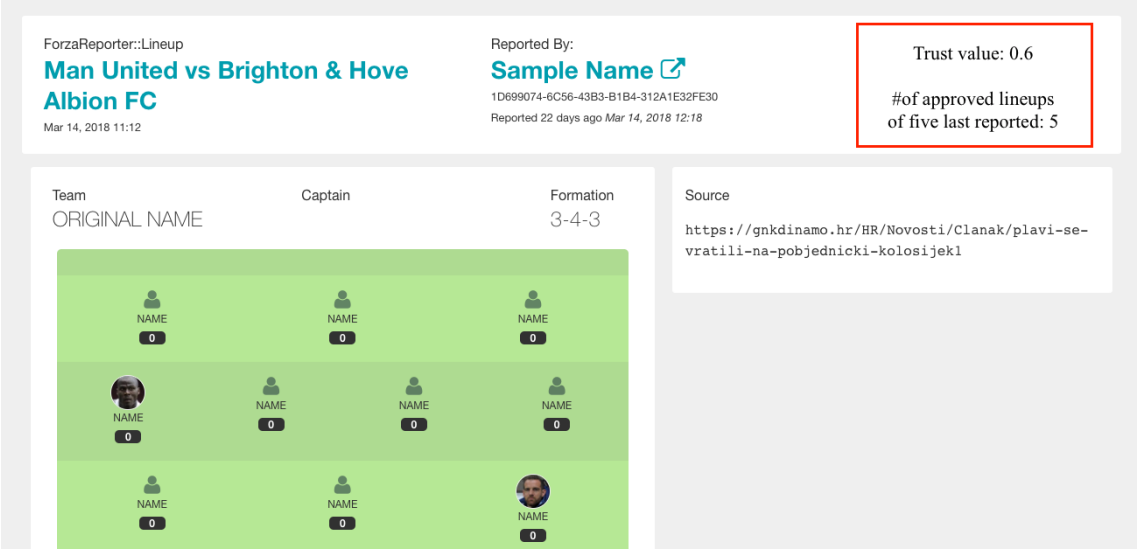


Figure C.29: An example of a user with trust value: 0.6, #approved lineups out of the last five: 5.

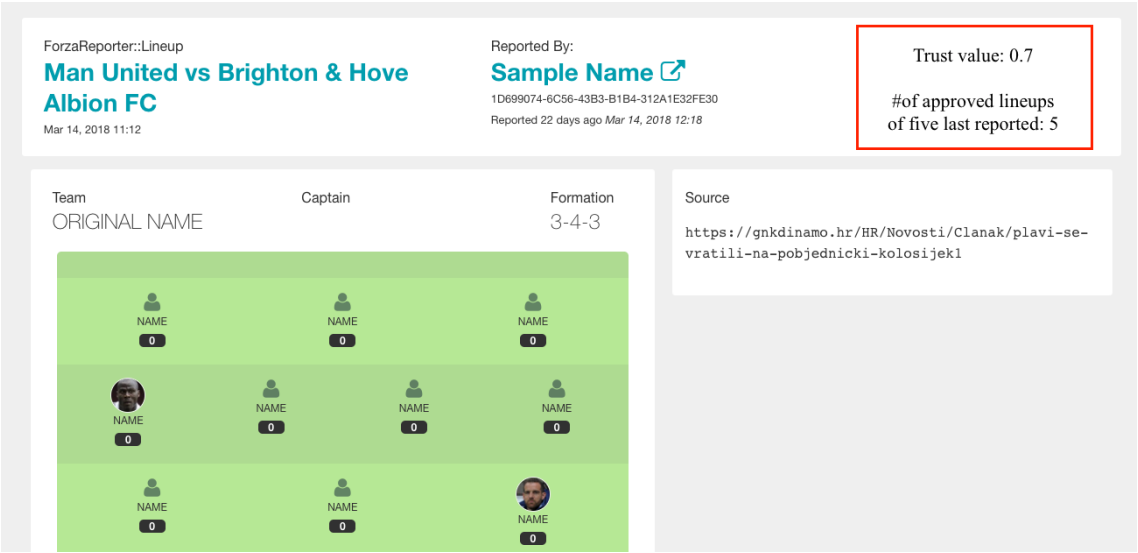


Figure C.30: An example of a user with trust value: 0.7, #approved lineups out of the last five: 5.

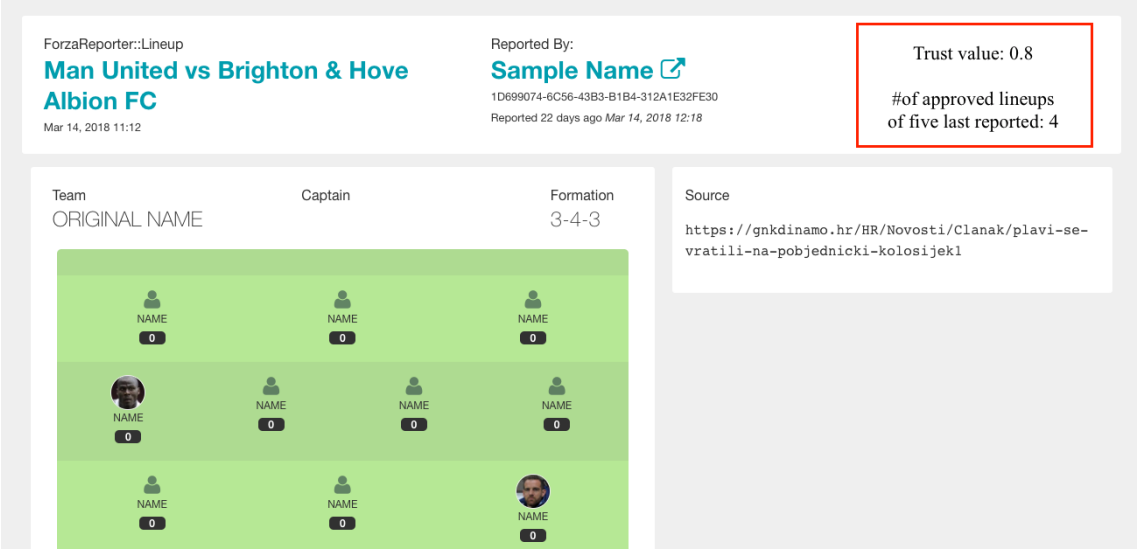


Figure C.31: An example of a user with trust value: 0.8, #approved lineups out of the last five: 4.

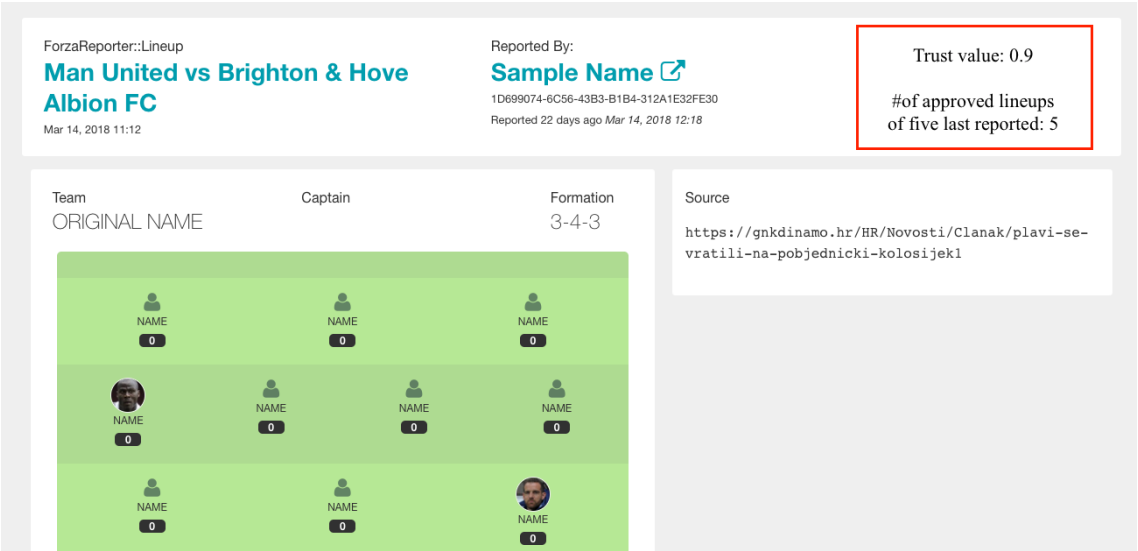


Figure C.32: An example of a user with trust value: 0.9, #approved lineups out of the last five: 5.

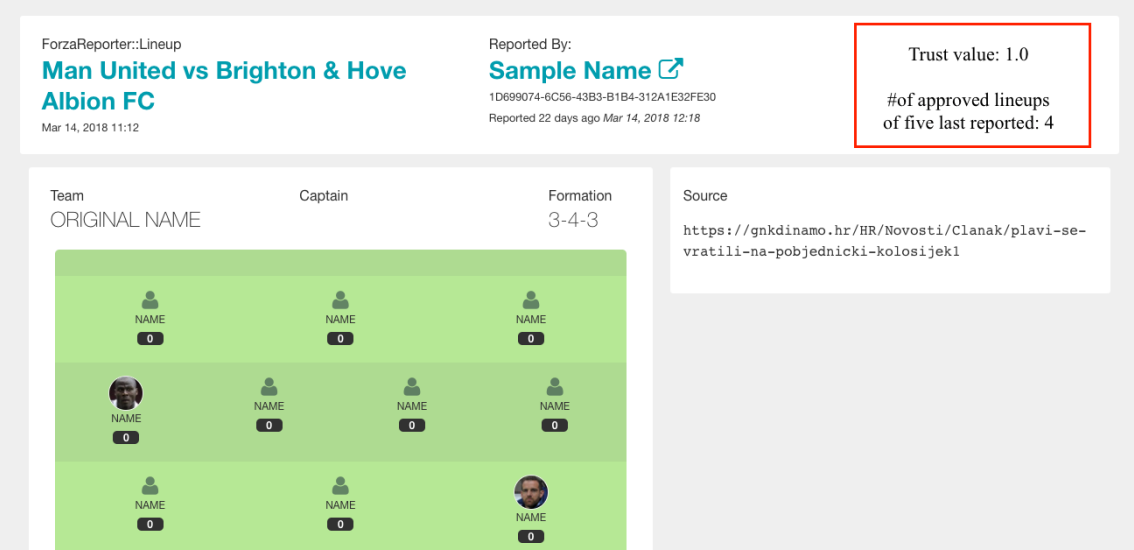


Figure C.33: An example of a user with trust value: 1.0, #approved lineups out of the last five: 4.

C.4 Users' approval rate together with number of reported lineups.

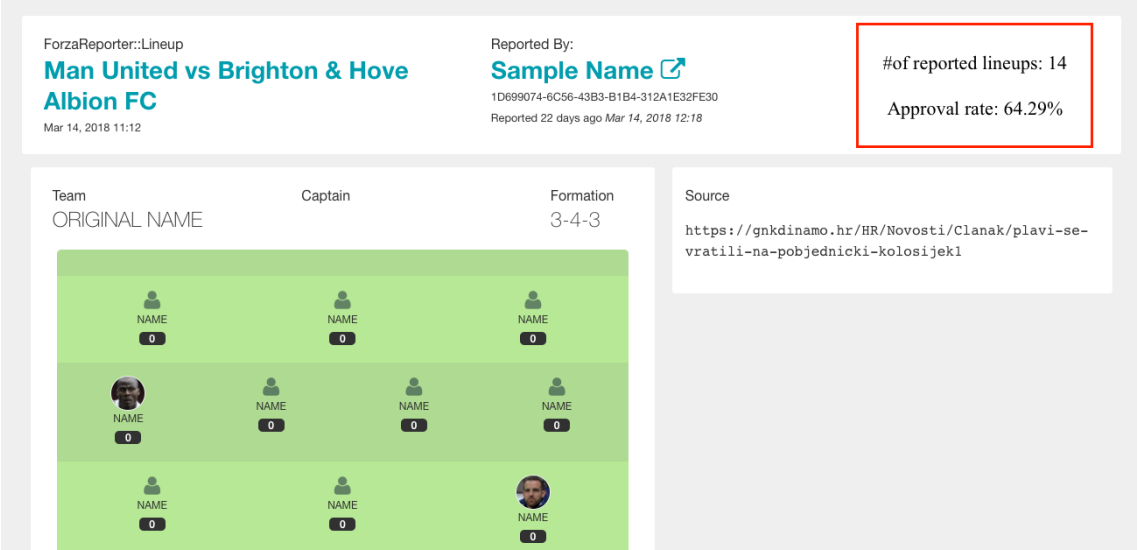


Figure C.34: An example of a user with approval rate: 64.29%, #reported lineups: 14.

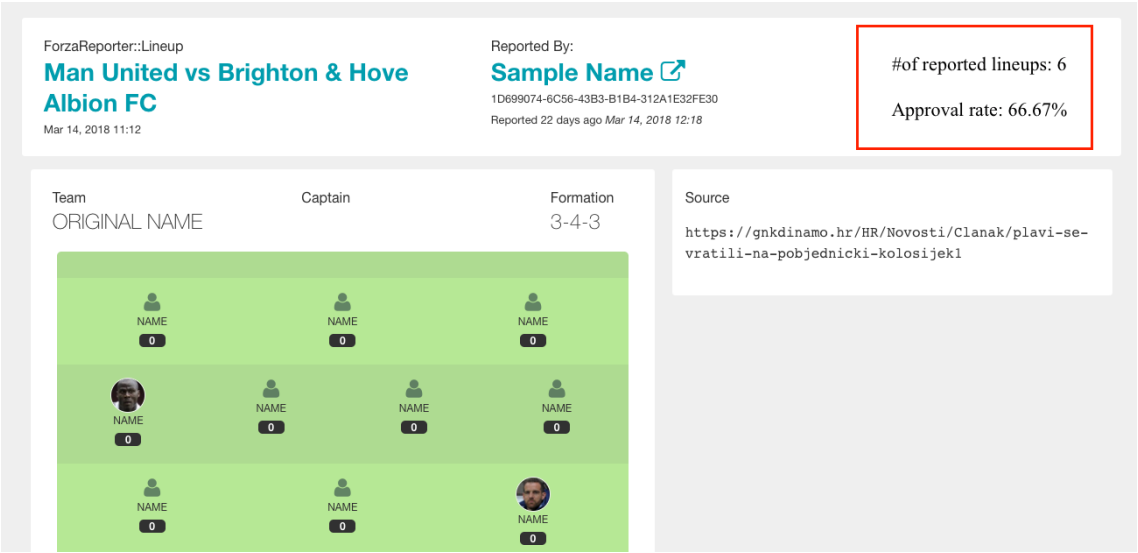


Figure C.35: An example of a user with approval rate: 66.67%, #reported lineups: 6.

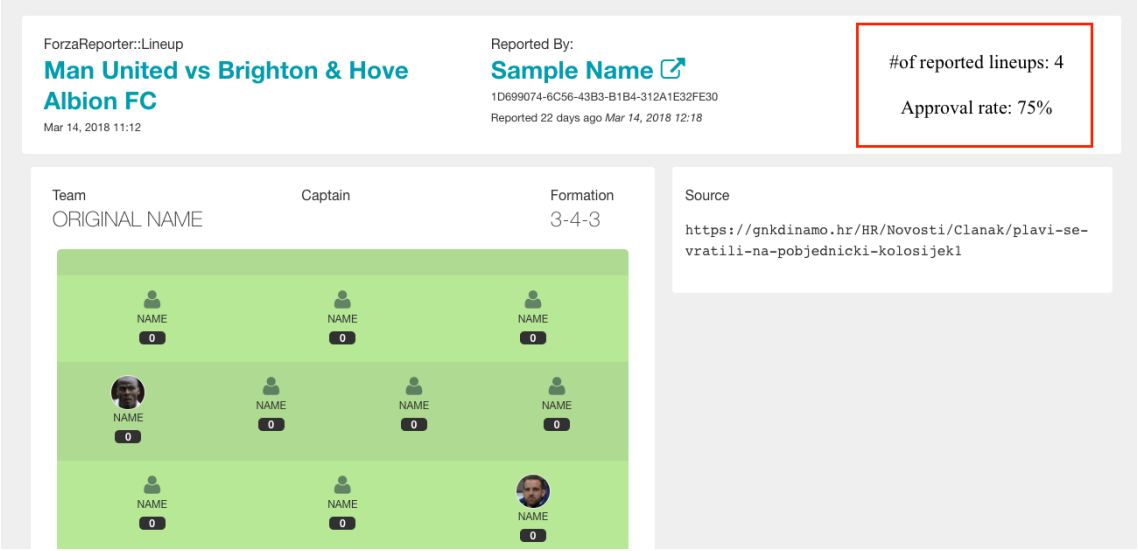


Figure C.36: An example of a user with approval rate: 75%, #reported lineups: 4.

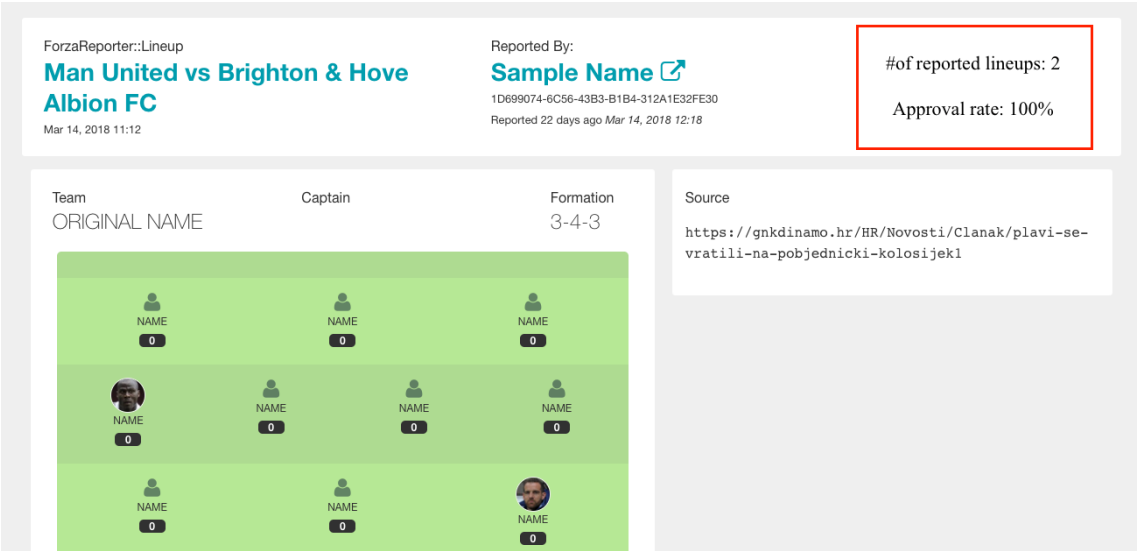


Figure C.37: An example of a user with approval rate: 100%, #reported lineups: 2.

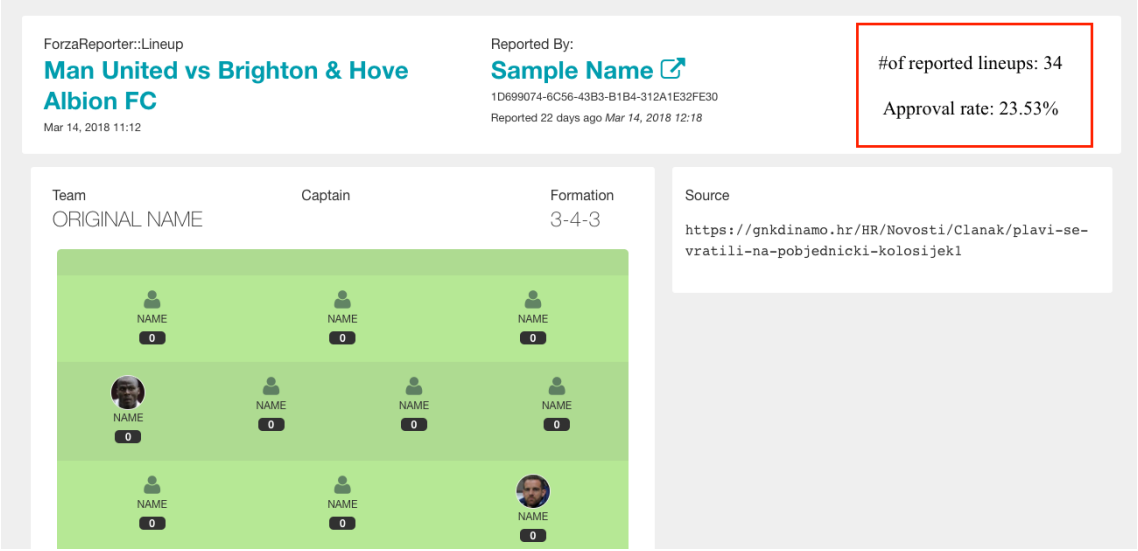


Figure C.38: An example of a user with approval rate: 23.53%, #reported lineups: 34.

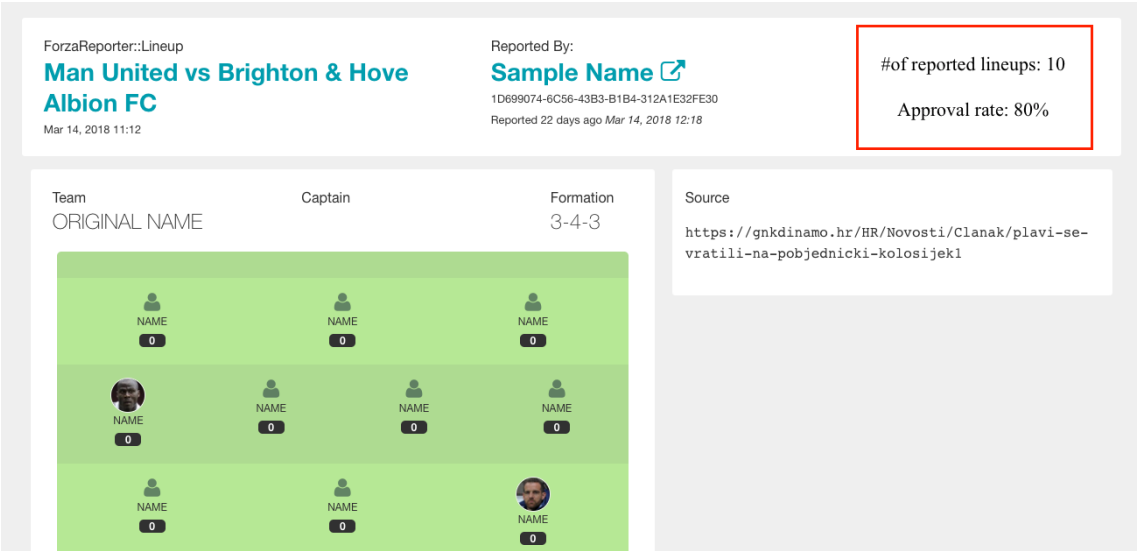


Figure C.39: An example of a user with approval rate: 80%, #reported lineups: 10.

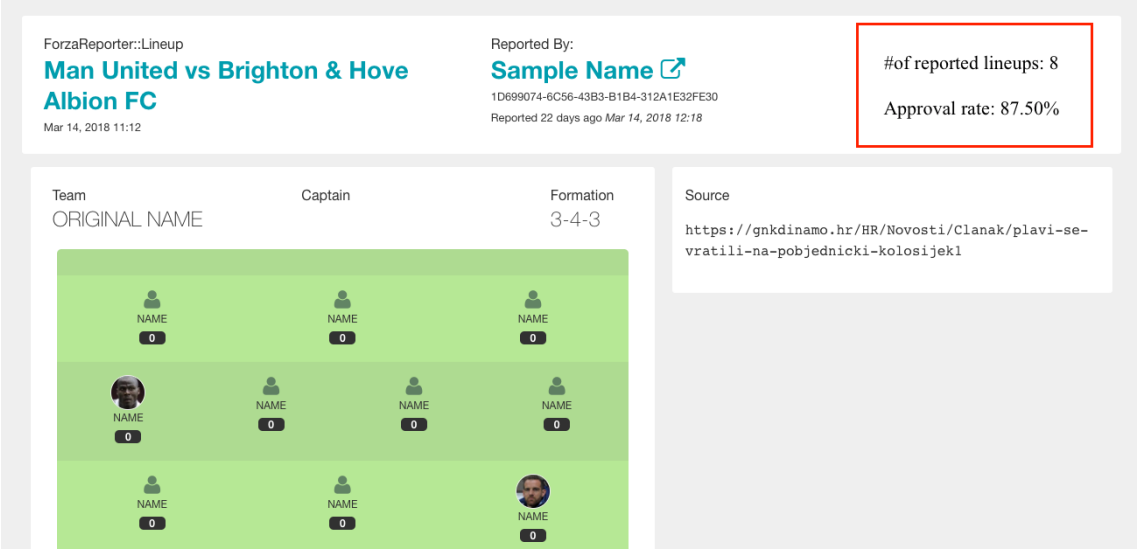


Figure C.40: An example of a user with approval rate: 87.50%, #reported lineups: 8.

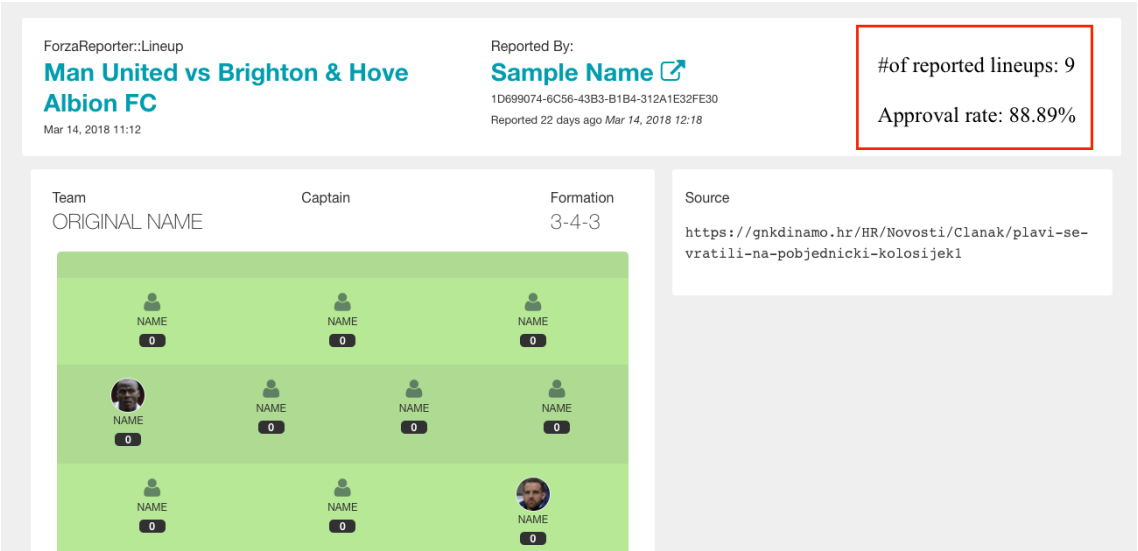


Figure C.41: An example of a user with approval rate: 88.89%, #reported lineups: 9.

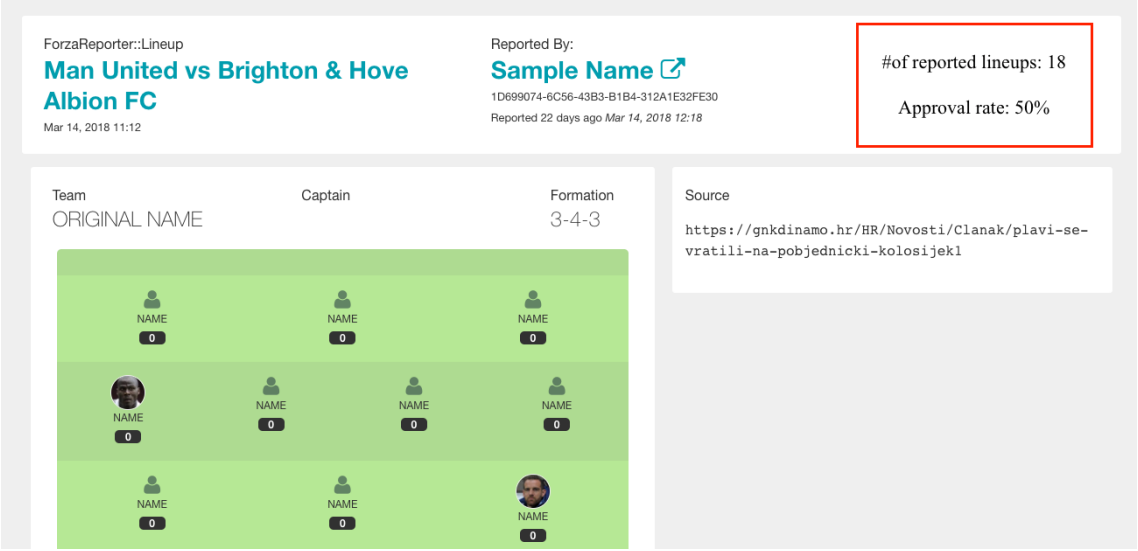


Figure C.42: An example of a user with approval rate: 50%, #reported lineups: 18.

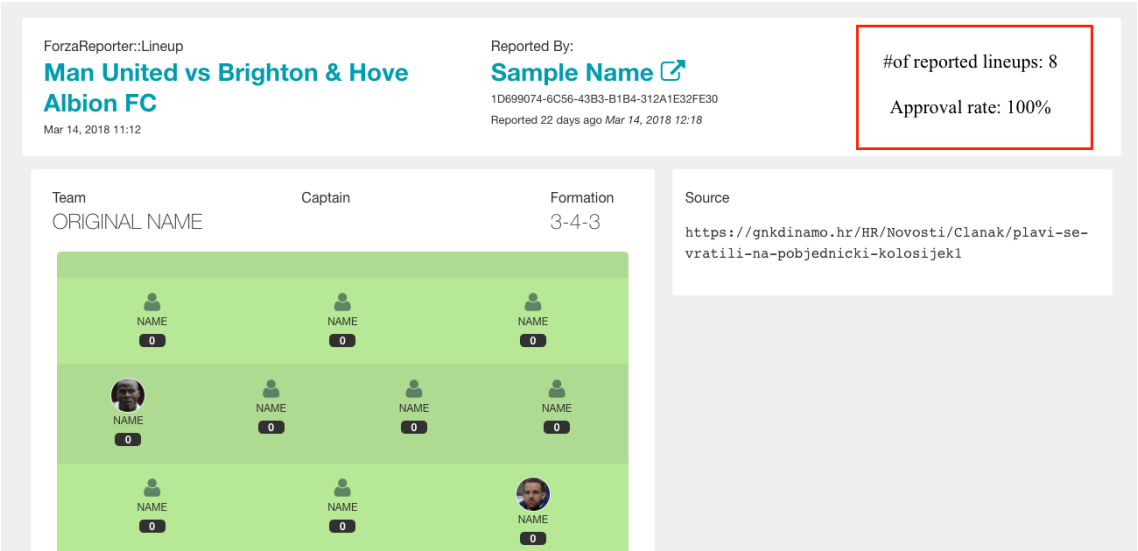


Figure C.43: An example of a user with approval rate: 100%, #reported lineups: 8.

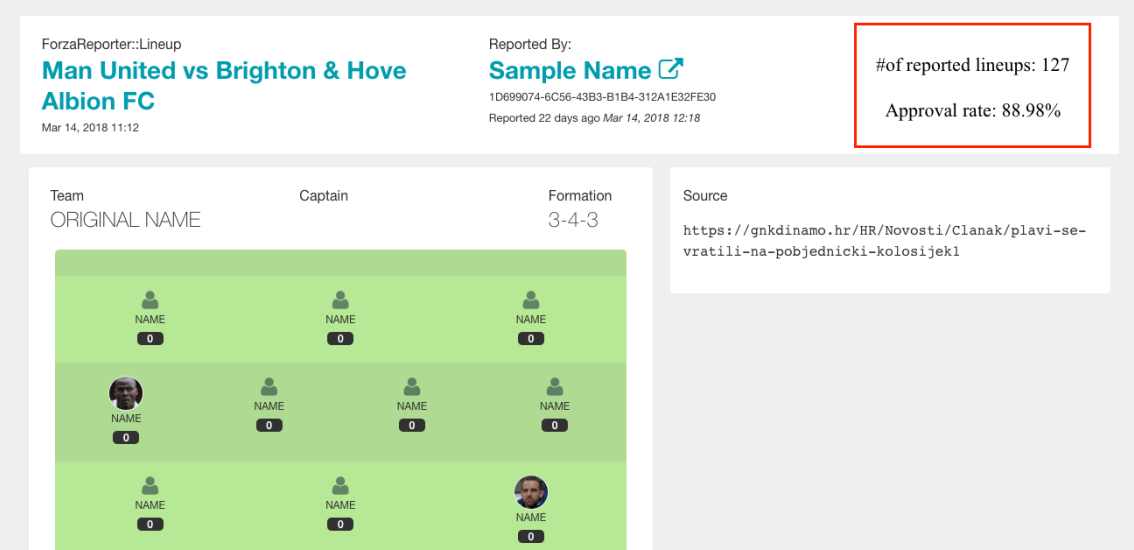


Figure C.44: An example of a user with approval rate: 88.98%, #reported lineups: 127.

D

Material used in the validation sessions for what feedback should be provided to users.

In this appendix, all material used in the fourth iteration is presented. In the forthcoming section, the drafts of user profiles that was used as a basis for asking the participants questions is displayed. Thereafter follows a section including the questions that were used in the open-ended interviews conducted in order to validate what kind of feedback is more suitable to provide to users.

D.1 Drafts of user profiles

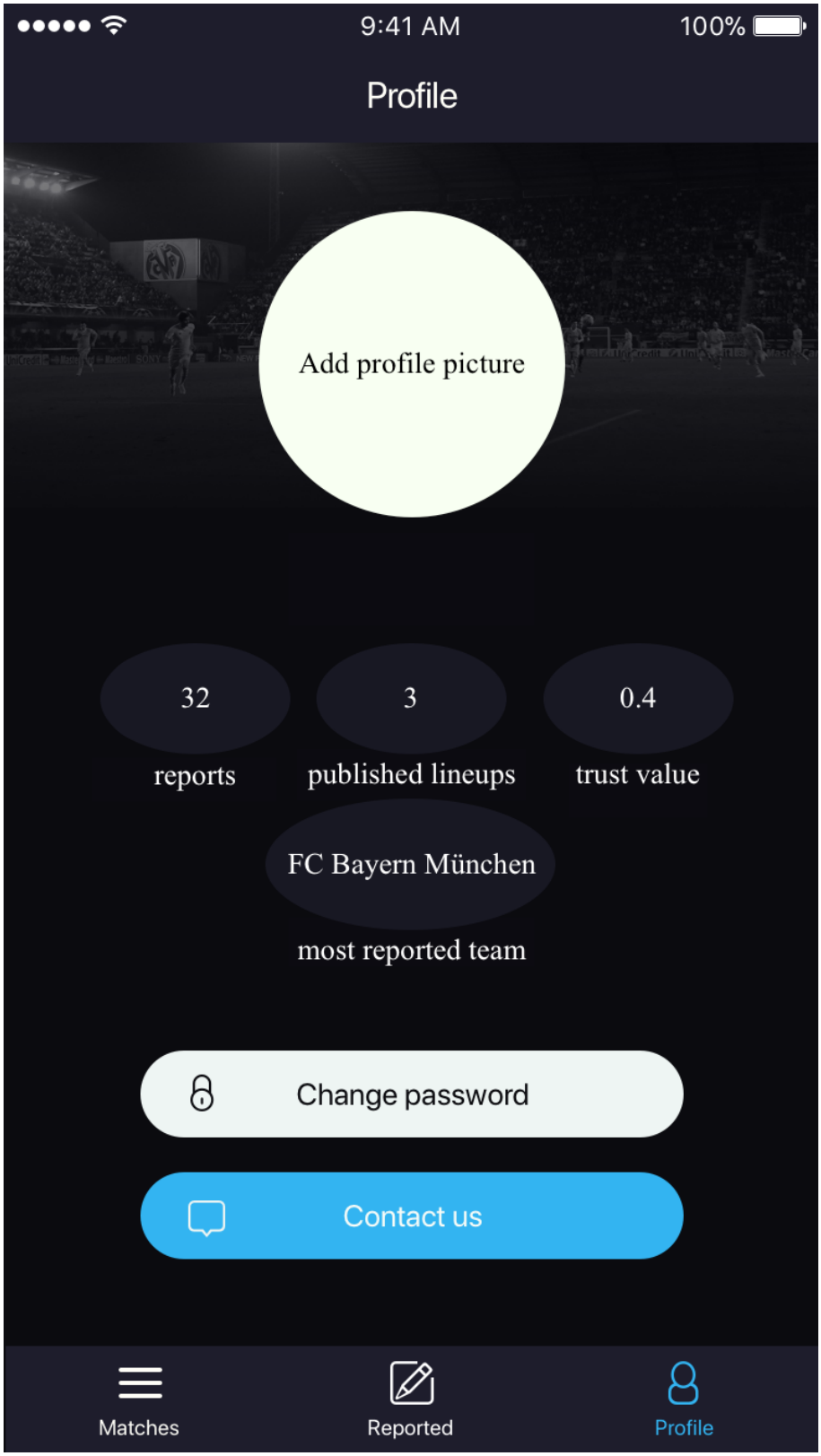


Figure D.1: Draft of user profile with trust value.

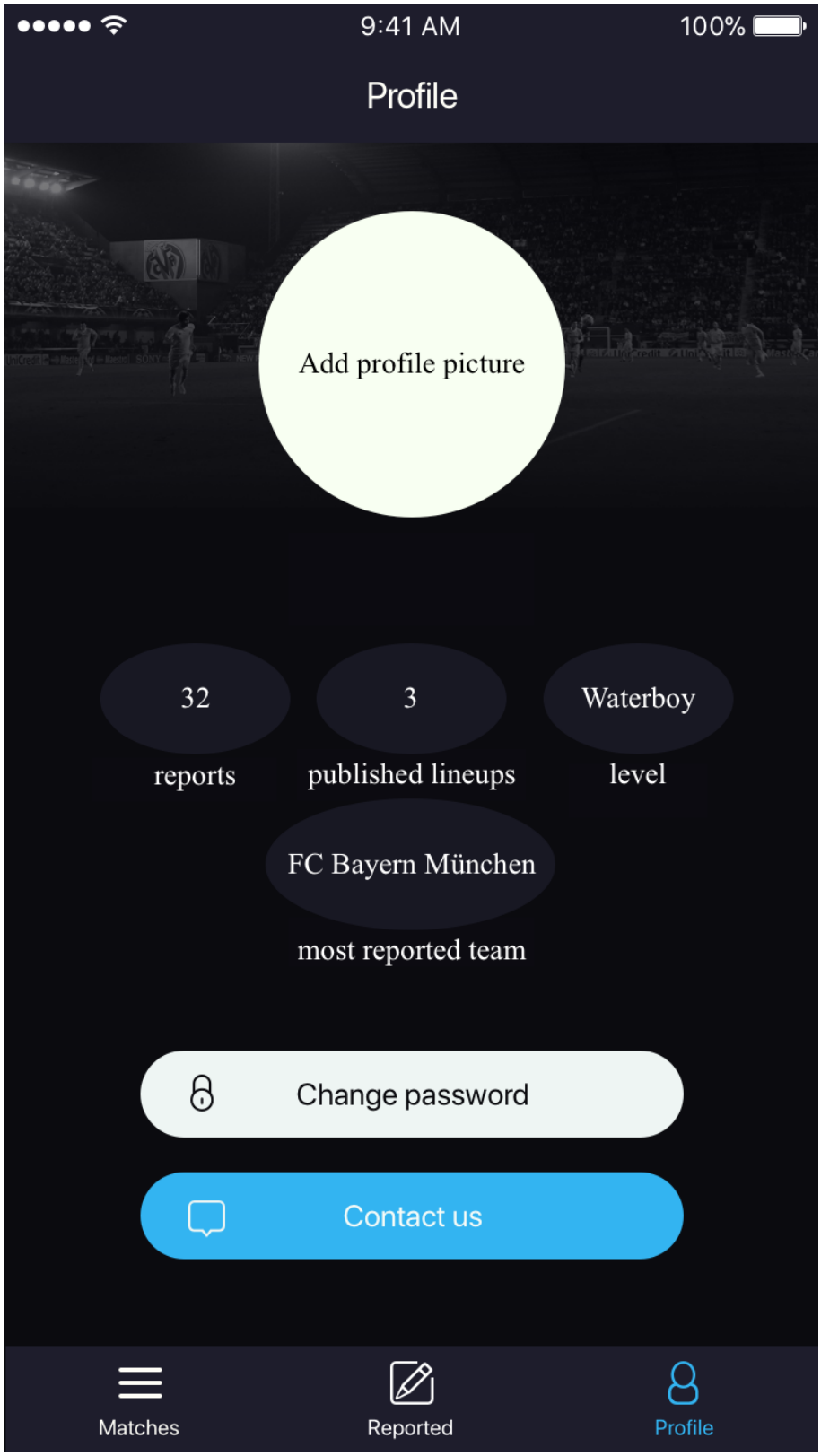


Figure D.2: Draft of user profile with a gamification element.

D.2 Questions for open-ended interview

In this part if the appendix, the questions that were asked to the participants in the open-ended interview conducted in conjunction to showing the drafts in D.1 are presented.

Trust value

- Do you think that the trust value should be displayed to the user? Why/why not?
- Would you as a user understand what that number means? If no, what further information would you like? Or is this not significant to you as a user?
- Do you think that the user should receive some kind of feedback regarding its performances? If so, what kind of feedback do you think is of significance?
- For each rejected lineup, should the feedback include the new trust value, the reason for it being lowered, and an explanation for what the user has to do in order to reach a higher trust value again, or only the reason for rejection?
- For each correct lineup, should the feedback include the new trust value and an explanation about what that new value means, or only the reason for the lineup being correct/approved?

Gamified element

- Should the trust value be gamified? Meaning, if a user for example has a trust value of 0.1, the user is a water-boy. Or if the user has 1.0 in trust value, the user is a coach?
- Should the trust value be displayed together with its gamification to the user? Or only the gamified value?
- Should the gamified value be reflected from the trust value or from a different value? For example, the number of reported lineups? The number of approved lineups in a row? Or any other value that you can think of?
- Should the user receive feedback about what each gamified value means and what could that feedback be?
- Should the user receive feedback about all the gamified values that exist and what the user needs to do in order to reach the next value in the hierarchy? If not, do you have another suggestions?
- Is there any feedback that you are missing? That you think would be of value to include?