universität wien

# MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„The joint influence of Task Difficulty and Performance Feedback on Advice Utilization, Judgment Accuracy and Overconfidence“


verfasst von / submitted by

Julia Wulf, B.Sc.


angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2020 / Vienna 2020

| | |
|---|---|
| Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet: | UA 066 914 |
| Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet: | Masterstudium Internationale Betriebswirtschaft |
| Betreut von / Supervisor: | Ass.-Prof. Steffen Keck, PhD |

**Abstract**

The following paper takes a deeper look on the effect of task difficulty and performance feedback on advice taking, judgment accuracy and overconfidence in the advice taking process. It aims to answer the research question whether judges learn to use advice more optimal over time when they receive performance feedback. Three hypotheses were stablished to test for the progress of advice taking, judgment accuracy and the three types of overconfidence. 103 participants had to estimate the kilocalories of food and provide information on their confidence level. The results showed that advice taking was low and also participants rather showed signs of underestimation instead of overestimation. Yet, judgment accuracy increased over the rounds as participants learned to state more accurate decisions and to adjust their 90% confidence intervals which led to lower overprecision. With regard to overplacement, no statement could be made as the results were too fluctuating over the rounds.

**Content**

## 1. Introduction

Despite research on the subject of advice giving and taking has been done for about twenty years (Brehmer & Hagafors, 1986), people often believe that the decisions they make are made independently or on their own. In fact, most decisions are made under influence as humans are known to be social beings. Whether how and in which stock to invest or which job to take people tend to consult others in important decisions as advice seeking and utilization represent fundamental tools before making decisions (Yaniv, 2004a; Yaniv & Milyavsky, 2007). People strive for social and cognitive interaction to represent their decision problem in a better way (Yates et al., 1996; Zarnoth & Sniezek, 1997) and to achieve an improvement of their final choice (Gino et al., 2009; Harvey & Fischer, 1997; Yaniv et al., 2011).

A very common phenomenon in the research of advice taking is the so-called Judge-Advisor System (JAS, see Sniezek & Buckley, 1995; Sniezek & Van Swol, 2001). To set the basis of a proper understanding of the following paper it is necessary to have a closer look on the used terminology of the JAS. This system implies the *judge* as the decision maker that receives the advice and has the opportunity to use or discount it as well as one or more *advisor(s)* as the ones who give recommendations on the particular task before the judge is making the final decision (Harvey & Fischer, 1997; Yaniv, 2004a). Moreover, decision tasks can be of either a *choice type* where the judge is presented with multiple options or a *judgement type* where quantitative estimates are required (Bonaccio & Dalal, 2006). While in the first case the advice appears as a certain recommendation of the advisor for a particular alternative (Sniezek & Buckley, 1995), in the second case the advisor provides the judge with an own estimation (Yaniv, 2004a, 2004b). It is noteworthy that in the traditional JAS the judge is provided with the advice disregarded the judge's decision whether or not to receive advice (Gino & Moore, 2007).

Past research has shown that in general people underweight the advice they receive (Yaniv, 2004b; Yaniv & Kleinberger, 2000). Harvey and Fischer (1997) even found that in cases where the judge's

initial estimation conforms with the advice such that the weight distribution for each estimation would be 50%, the actual weight of advice (WOA) would be around 20% to 30% regarding the final estimate. On the other side there is also evidence about overweighting advice. For example, companies would hire business management consultants and pay large amounts of money to receive advice on business issues although they won't know enough about the value of the consultancy's recommendation (Micklethwait & Wooldridge, 1996). Moreover, the research area of judgment revision after advice receipt has found high interest among researchers (Harvey & Fischer, 1997; Yaniv, 2004a) and led to essential findings such as that judges often underweight the advice they receive from others which results in egocentric discounting (Yaniv & Kleinberger, 2000). Although the combination of judge's and advisor's estimates has been found to be a useful technique (Armstrong, 2001; Larrick & Soll, 2006), in only the fewest cases averaging is actually used (Harvey & Fischer, 1997; Soll & Larrick, 2009; Soll & Mannes, 2011; Yaniv & Choshen-Hillel, 2012).

Besides the characteristics of judge and advisor, advice taking depicts the weight and integration of the judge's initial choice with the advice given by the advisor. This leads to advice taking being an essential kind of information processing (Brooks et al., 2015). In addition to the fact that a combination of independent opinions received leads to an increase in the judge's accuracy (Harvey & Fischer, 1997; Yaniv, 2004a, 2004b), the aggregation of multiple recommendations as well yields higher accuracy due to the adjustment of random errors (Galton, 1907; Herzog & Hertwig, 2009). Unfortunately, the area of research regarding performance feedback, especially in relation to advice taking is scarce. Regarding the field of management, Min et al. (2020) underline that performance information has to be used to have a positive impact on performance. Other researchers showed that receiving performance feedback in a constant manner can help to enhance the participants' learning goal orientation because subjects succumb to the norm of striving for learning progress (Ghoshal & Bartlett, 2007).

This paper aims to answer the research question whether participants (here in the roles of judges) learn to use advice in an optimal way over time when they receive performance feedback under consideration of different task difficulty levels. To test the stated hypotheses an online survey was presented to 183

participants where they had to estimate the amount of kilocalories of food shown in the pictures and answer some questions regarding their levels of overconfidence.

The present paper is structured as follows: beginning with a comprehensive literature review the most significant findings in the subject of the involved variables are presented. The first part comprises the different types of advice, the circumstances and measurements of advice utilization and advice discounting as well as other influencing factors on the advice utilization process. This is followed by two sections addressing judgment accuracy and the different types of overconfidence by referring to the corresponding literature. Afterwards, popular research on task difficulty and performance feedback is presented. This section ends with an overview of the research gap. The second part focuses on the conducted online study which is based on the papers of Gino and Moore (2007) as well as Yaniv and Choshen-Hillel (2012). After an overview of the stated hypotheses, the methodology is explained, giving information about procedure, participants and dependent measures. This is complemented with a proper description of the results that were achieved in the data analysis. In the final discussion, the most important findings and possible justifications are elaborated. The third part rounds the thesis off with a conclusion. Based on the research literature and the gained results limitations and suggestions on a future research approach are reflected.

## 2. Literature Review

Subsequently, terms for the agents of the JAS will be used. In this connection, the judge is also referred to as the decision maker or in some cases as the advice seeker while the advisor is also known as the advice giver. According to the literature, advice taking and advice utilization are used equivalently.

### 2.1. Advice

The term of advice can have many definitions that differ only in a tiny detail from one another. The Cambridge Dictionary (*Advice*, 2020) for example defines advice as '*an opinion that someone offers you about what you should do or how you should act in a particular situation'*. According to several researchers among which are also Bonaccio and Dalal (2006) the definition of advice can be widely displayed as it is not strictly defined what is considered an advice and what not. They further state that not only recommendations which they found to be associated with giving advice in the first place but also social or emotional support as well as the provision of guidance in the decision-making process can have a more or less strong impact on the final decision of the judge (Dalal & Bonaccio, 2010). These influencing types of advice will be presented and discussed in the following.

#### 2.1.1. Types of Advice

One meaning of advice that is most common in the research area of advice taking is the term *recommended for*. This implies the expression of the advisor's opinion, the person that is asked for advice, for a specific problem or question stated by the judge, the person seeking for or receiving the advice in order to make a decision (Bonaccio & Dalal, 2006; Dalal & Bonaccio, 2010). An opposing type of advice is indicated by the advisor's *recommendation against* one or several options for a specific problem or question stated by the judge. This type of advice occurs when the advisor expresses an opinion for the exclusion of an action, leaving the judge to choose from the options that are left. Although there seems to be rare sources in literature regarding this type of advice, it can be clearly compared to the findings of Bonaccio and Dalal (2006; Dalal & Bonaccio, 2010). The third type

comprises the effect of *social support* from the advisor for the judge. It occurs in situations where the advice seeker is in need of social-emotional assistance for a specific problem or question. The authors define this phenomenon rather as a form of interpersonal support than a typical type of advice (Dalal & Bonaccio, 2010). The so-called *decision support* sets the fourth type of advice. Here, the advisor does not take the position for or against an opinion as in the other types but rather implies support in the judge's decision process where the focus is rather on the methodology of decision making than on the result of the decision itself (Dalal & Bonaccio, 2010). The phenomenon of spreading *information* can be considered as the fifth type of advice. This simply implies the provision of additional information by the advisor that might be helpful to the judge. As in the case of the fourth type, providing information is independent from advancing a view (Dalal & Bonaccio, 2010). It's interesting to see that Dalal and Bonaccio (2010) did research on the question which type of advice is preferred by the judge when receiving assistance from an advisor. After conducting two studies they showed that the *information* type of advice was the most favored one. Although types like *recommendation against* and *social support* were positively evaluated as well, the researchers found that receiving information about the possible alternatives of the advice giver when making a decision seemed more useful to the decision maker. Nonetheless, a recommendation as advice is favored in situations where the judge does not ascribe high importance to the own autonomy or where the advice giver is considered to be reliable.

Further research done by Sniezek and Buckley (1995) comprises the definition of the advisors as the ones who *'formulate judgements or recommend alternatives and communicate these to the person in the role of the judge'* (p. 159). A specific form of advice (i.e. *'choose option X'*) was found to be of the least favored types of advice by decision makers as these recommendations might not always meet the judge's objective for the advising interaction (Horowitz et al., 2001). Yet others (Goldsmith & Fitch, 1997; Horowitz et al., 2001; Whittemore et al., 2000) go further and define the advisor as someone who provides not only task-related suggestions but also socio-emotional support, assistance in problem-solving and suggestions regarding specific approaches. Moreover, recommendations against one or more of the options available can be added in. Evidence was found that advice seeking not only exists because of the judge's perception of receiving unknown information but also can help to reduce the

judgmental decision error and hence, to achieve a better judgement performance as well as to boost their confidence in the decision task but also due to expected social support (Heath & Gonzalez, 1995). Other researchers argue that it is difficult to properly analyze the process between judge and advisor within advice giving and taking. They therefore point out five types of advice which are: *solutions, meta-knowledge, problem reformulation, validation, and legitimization* (Cross et al., 2001). They further come up with the findings that those types of advice are all located on a sole dimension in the presented order such that an advisor giving one type of advice would also be able to provide the types lying below. In addition to this conclusion, they further speculate that advisors tend to provide the first type of advice when asked for, and after that the other, less precise, types. In contrast to that, Gibbons (2003) proposed the inclusion of other factors in the spectrum of types of advice such as emotional support, support of judge's initial decision, information or argumentation, suggestion of new options, and assistance either towards the decision process or for the self-insight of the judge.

### 2.1.2. Reasons for Advice Utilization

Some reasons for decision makers to ask for and use advice are, amongst others, the sharing of responsibility for the decision's outcome and the attempt to increase the probability of more accurate or even optimal judgments (Harvey & Fischer, 1997; Yaniv, 2004a, 2004b). Indeed, according to the literature, the most common reason for taking advice is to enhance the own accuracy. Yaniv (2004b) showed in studies involving forecasts of economic events and estimations about historical data that a significant improvement of final decisions can already be achieved with only three to six answers or recommendations from the advisor(s). By conducting a study concerning the influence of advice from others, the author found that although the involved five experts independently achieved correct results in 65% of the time, a combination of the stated opinions would have achieved a score of 76% (Yaniv, 2004b). Psychological and statistical factors affect the level of accuracy improvement also due to the existence of a random error. Thus, whereas the regard of only one advice source can induce a higher random error, the joint consideration of multiple answers can lead to a reduction of it and thus, to a higher decision accuracy (Herzog & Hertwig, 2009; Yaniv & Choshen-Hillel, 2012). Another study

revealed that accuracy improvement in estimation tasks could be increased by approximately 20% in cases where participants considered the opinion they received from an advisor. At this point, a crucial factor to be kept in mind is that the advisor does not necessarily possess a higher degree of knowledge or intelligence compared to the decision maker. A more significant criterion is the origin of the received advice, namely that these come from independent sources (Yaniv, 2004a). Despite the constant presence of a certain dependence between the advisors, advantages in terms of accuracy are also gained when the received advice is only moderately or even low positively correlated (Johnson et al., 2001; Yaniv, 2004b).

Another reason that is justifying advice taking is found in the arising psychological benefits. Individuals are expected to seek for advice more when the decision is of specific importance or the decision maker feels a strong responsibility for the outcome which can also result in advice seeking in order to share this responsibility with others (Harvey & Fischer, 1997; Yaniv, 2004a). Social benefits are also represented in the advice taking process. In this case, decision makers intend to achieve explanations for their own actions by seeking advice in terms of self-presentation (Kennedy et al., 1997; Yaniv, 2004a). Besides the psychological and social advantages aimed for by the decision maker, Schotter (2003) as well as Heath and Gonzales (1995) argue that advice taking also enables the decision maker to change the perspective by involving others' viewpoints and consider new information. In other words, taking on another perspective changes and enhances the own structure of thinking. Besides, also emotions can have a certain influence on the process of advice utilization even if the feelings are not directed towards the advisor. For example, anger can elicit advice discounting whereas the feeling of gratitude enhances the judge's susceptibility to advice (Gino & Schweitzer, 2008). Other researchers confirm this argumentation in a sense as positive emotions directed towards others, like gratitude, can in fact enhance the decision maker's susceptibility to advice due to the resulting judge's perception of the two agent's motivations matching. In turn, emotions of a negative state directed towards others, like anger, can result in the opposite effect (de Hooge et al., 2014). Contrary to that, positive emotions directed towards oneself, like pride, can decrease the judge's susceptibility to advice while negative emotions, like shame, are likely to enhance the susceptibility. Also, the feeling of anxiety affects the

judge's confidence as it decreases it which, in turn, strengthens the decision maker's susceptibility to advice (Gino et al., 2012). Gino and Schweitzer (2008) as well focused on the emotions of a judge or, more precisely, on anger and gratitude. They found that incidental emotions which are emotions that are formed by a situation-irrelevant experience made beforehand had a significant impact on advice taking. It was shown that grateful participants were more willing to take advice than participants with neutral emotions who, in turn, had more trust in the received advice than individuals who felt anger. This phenomenon was experienced by many participants and led to the conclusion that trust serves as a mediator on the relation between the participants' emotions and their advice taking behavior such that negative emotions decrease the trust towards the advice. Furthermore, the accuracy in final decisions increased for people exhibiting higher trust levels towards the advice due to their positive attitude of emotions.

With regard to an improved use of advice, researchers found that the optimal strategy for advice utilization that yields the most accurate results is averaging. Soll and Larrick (2009) illustrate the advantages of this strategy by presenting an example including the estimation of future inventory levels according to the forecasted demand for a stored product. Imagine the correct value of demand is 70 pieces. The advisors' suggestions can either be on the same side of the truth (i.e. take the value of 50 or 60) or bracket the truth (i.e. take the value of 60 or 90). In the case of the recommendations being on the same side of the true value, the first advisor is falling short of the true value by 20 pieces and the second advisor by 10 pieces, resulting in an average missing of 15 pieces. Calculation with the averaging technique would yield 55 pieces which falls short of the true value by 15 pieces. In the case of bracketing the truth, both advisors miss the correct answer by the exact same values, namely 10 and 20. Here too, the average value of missing the correct answer is 15 but the difference of this case to the former one is that the averaging technique would result here in a value of 75 which is much closer to the correct answer than 55 in the first case. Minson et al. (2011) found that the benefit of advice is the highest when the recommendation of the advisor does not support the opinion of the judge but rather takes on a value on the other extreme of the correct answer compared to the judge's initial opinion. Furthermore, in the absence of information addressing the quality of advice, techniques such as averaging the values of the

opinions or choosing the median is helpful to achieve a final judgment as accurate as possible (Minson et al., 2011; Yaniv & Milyavsky, 2007). Nevertheless, Yaniv (1997) realized that when judges have access to information regarding the quality and correctness of the received advice, techniques like weighting the information represented more advantageous strategies than averaging or trimming. Other researchers suggest that a more optimal use of advice can be achieved when the sources of advice are distinguishable from each other (Harvey et al., 2000). In their study, participants in the role of decision makers were able to use advice better when they received advice from multiple advisors who provided prediction trends that were not usual and extraordinary for the task which made those advisors standing out.

### 2.1.3. Reasons for Advice Discounting

While advice taking (or advice utilization as these terms are often used equivalently) describes the extent to which the judge as decision maker takes the given advice into account, advice discounting refers to the opposite, namely refusing to take the advice (Bonaccio & Dalal, 2006). The decision for taking or discounting the advice can, amongst others, be influenced by the advice distance, meaning the difference between the judge's initial judgment and the advisor's recommendation. The greater the distance the more likely advice discounting will occur (Yaniv, 2004a; Yaniv & Milyavsky, 2007). Interestingly it was found that this event takes place mainly for judges with high knowledge. Moreover, the more different the suggestions of one advisor are from those of the other advisors the more likely judges will apply advice discounting (Harries et al., 2004). Other researchers found that advice discounting occurs less in situations where the advice comes from advisors that are perceived to have high expertise (Goldsmith & Fitch, 1997; Harvey & Fischer, 1997; Sniezek et al., 2004) or seem to be wiser (Feng & MacGeorge, 2006) and that a higher task complexity also supports advice discounting (Gino & Moore, 2007). Additionally, advice discounting can also be linked to the judge's perception of potential costs and benefits of the given advice (Schrah et al., 2006). On the one hand, according to Bonaccio and Dalal (2006), the benefits of advice taking comprise a greater judgment accuracy as well as shared responsibility for the results of the decision. On the other hand, greater effort such as some kind of

financial expenditure represent the costs. Advice discounting happens when the decision maker perceives the costs as greater than the benefits.

As already mentioned most individuals rely on their own opinion more than on the advice they receive (Gardner & Berry, 1995; Harvey & Fischer, 1997; Yaniv & Kleinberger, 2000). Thus, instead of following the advice which would bring better results they tend to discount it (Bonaccio & Dalal, 2006; Yaniv, 2004b; Yaniv & Kleinberger, 2000). There are researchers who argue that the reasons for advice discounting are the difference of information (Yaniv, 2004a, 2004b; Yaniv & Kleinberger, 2000), the occurrence of the anchoring effect (Tversky & Kahneman, 1974) and the appearance of an egocentric bias (Krueger, 2003). In the case of the first phenomenon, namely the difference of existing information, the judge has access to the own internal justifications that create the personal opinion disregarded the internal justifications of the advisor(s) (Yaniv, 2004a; Yaniv & Kleinberger, 2000). The second reason mentioned, the anchoring effect, implies that the judge's initial opinion serves as a point of reference, modified when advice is received. But respective to the advice taking process, this type is deficient as it leads to egocentric discounting, the third phenomenon (Harvey & Fischer, 1997; Lim & O'Connor, 1995). This, in turn, gets decision makers to choose the own opinion justified by their beliefs of being superior over the advice given by others (Harvey & Fischer, 1997; Krueger, 2003; Yaniv & Kleinberger, 2000). Harvey and Harries (2004) emphasize the difference between egocentrism and anchoring by defining the first one as a long-term influence on the judge's opinion while anchoring rather represents an evaluation of a new situation and hence, a short-term influence. In addition, the phenomena of the anchoring bias or a certain adjustment strategy can also result in advice discounting (Tversky & Kahneman, 1974). In these cases the judge considers the own initial decision as the anchor that is adapted according to the advisor's given proposal (Harvey & Fischer, 1997; Lim & O'Connor, 1995). To dig deeper into the aforementioned reasons for advice discounting, one finding in the JAS research is considered to be of the most resilient ones which is the *egocentric advice discounting* (Yaniv, 2004a; Yaniv & Kleinberger, 2000) which describes the judge's tendency to overestimate the own opinion compared to the advisor's one (Harvey & Fischer, 1997; Krueger, 2003; Yaniv & Kleinberger, 2000). Although using advice would help to improve the judgment accuracy, researchers found that decision

makers often only move around 20% to 30% towards the advice they had been given as they think of themselves to perform better (Harvey & Fischer, 1997; Krueger, 2003). It was also found that judges would exhibit this bias even when being confronted with new problems or situations as well as when receiving advice prior to making a decision. Just as mentioned above, Yaniv and Kleinberger (2000) and Yaniv (2004a, 2004b) state that the reasons for the occurrence of advice discounting lie in the judge's facilitated approach to the own internal justifications and evidences the decisions are based on while the decision maker lack knowledge about the advisor's understanding about the given problem.

Additional research regarding the *egocentric bias* found that judges tend to ascribe more value to their own decisions than to the received opinions from others. This overweighing of one's own opinion also results in inaccurate decisions. According to researchers, on the one hand the occurrence of egocentric advice discounting is caused by the judge's perception of the own knowledge and decisions to be predominant to the opinions of others. On the other hand, the judges have insight into their own justifications and the thinking process that led them to the stated decision. However, people seldom have access to the internal reasoning of others that justifies their advice and hence, this creates a certain knowledge difference that is likely to cause an enhanced judge's overweighting of the own opinion (Yaniv, 2004b; Yaniv & Kleinberger, 2000). In relation to that, the absence of the opportunity for judge and advisor to communicate with each other and to discuss a consensus prior to the formation of an independent and individual judgment can increase advice discounting. This was justified by the judge not being able to access the advisor's internal reasoning (Minson et al., 2011). . Interestingly, Minson and Mueller (2012) found that judges represented by a group consisting of two individuals are less likely to use the received advice compared to single judges. This is due to the fact that these so-called dyads exhibit a high activity in collaboration that increases the confidence and decreases the appraisal of external information which leads to the phenomenon of egocentric discounting (still see Schultze et al., 2013 for the response to Minson and Mueller and a different approach to the data analysis). The conclusion of these statements is that egocentric advice discounting occurs mostly in cases where the advisor's opinion differs from that of the judge (Yaniv, 2004a, 2004b; Yaniv & Milyavsky, 2007).

Other research on the heuristics affecting advice discounting found the *confirmation bias* to be a crucial one. Further research on this bias revealed that in many cases individuals discount advice as they don't feel that the information received does assist their initial opinion**.** Confirmation bias occurs in cases where independent individuals have been able to form a decision before receiving additional suggestions. Due to this already created pre-advice decision participants are more likely to discount the received advice as it appears to be not supporting their independent decision. Rather than that, independent judges were found to maintain their initial decisions even in events where they encounter additional information that proofs their own decision to be flawed (Ross et al., 1975). In addition to these findings, Yaniv and Choshen-Hillel (2012) argue that the creation of judgments prior to receiving advice is likely to enhance the egocentric bias and hinders the judgement accuracy of final decisions. Rader et al. (2015) introduce several terms to the topic of pre-advice and post-advice opinions. For example, the so-called *dependent advice sequence* represents the situation in which the advice precedes the initial opinion of the judge where, in contrast, the *independent-then revise sequence* describes the situation in which the judge has already made a decision before receiving the advice. In five studies, the researchers tried to examine the difference in the advice taking behavior in situations with a judge's pre-advice opinion and without and found that the final estimates of participants with a pre-advice opinion drifted towards the received advice and also achieved more accurate estimates while the estimations of participants without a pre-advice opinion drifted away from the received advice. This so-called *push-away effect* describes the lack of the subjects' confidence in the advice they get and leading judge's to consider alternatives. A contrary effect in this study was observed when the opinion distance between the judge's initial opinion and the advice was low as the confidence level rose. This leads to the finding that confidence serves as a mediator in the relationship between the advice sequence and the degree of the push-away effect.

Considering the quality of advice, the probability of a judge discounting inaccurate advice is higher than that of discounting accurate advice (although the probability for the latter is not zero: see Gardner & Berry, 1995; Lim & O'Connor, 1995; Yaniv & Kleinberger, 2000). As it is hard to gain a good reputation but easy to lose it due to inferior advice, judges tend to neglect poor advice (Yaniv & Kleinberger, 2000).

Moreover, the judge's perception of the quality of the advisor's past proposals can have an influential effect on the judge's advice taking as well and thus, on the judgement accuracy (Redd, 2002). The judge then applies knowledge about the advisor's past suggestions and further information about the forecasting strategy to make conclusions about the quality of the advice (Yates et al., 1996). Feng & MacGeorge (2006) found that more advice was taken from individuals that were perceived by the judge to have more expertise due to greater age, higher education, greater life experience or general wisdom compared to the decision maker.

Furthermore, research found that subjects tend to underestimate the quality of their own performance as well and think of achieving worse results than others when exposed to difficult tasks compared to easy tasks (Kruger, 1999; Moore & Kim, 2003; Windschitl et al., 2003). The reason for this event might be that subjects do not feel the need to consider recommendations from others when they believe to perform better than expected and thus, better than others. In contrast, when individuals believe to perform worse than expected and thus, worse than others the probability is higher that the judge might consider received recommendations in the decision-making process (Gino & Moore, 2007).

### 2.1.4. Measurement of Advice Utilization

There are numerous methods to measure advice utilization that can be applied on the Judge-Advisor system. As they differ with respect to the type of decision task, in general, choice tasks and judgment tasks are taken into account. Choices, for instance, represent decisions that include the possibility to select between options. For instance, a judge has to make the decision about which car to buy or which job to take in order to gain higher salary after 5 years. On the contrary, judgmental decisions are characterized by comprising estimates or forecasts related to measurable values. For example, a judge might see oneself confronted with a task to estimate how much money to invest in company shares or which country has the highest consumption of sugary beverages. The availability of choices is much higher in judgement tasks than in choice tasks as judgement decisions require data input consisting of

natural or ordinal numbers while nominal or categorical numbers are used in choice tasks (Bonaccio & Dalal, 2006). Measurements for these mentioned types will be discussed in the upcoming sub-sections.

### *Choice Tasks*

Thus, in most cases of choice tasks, advice taking is likely to be specified as *matching* which is a conformity between the judge's final opinion and the advisor's suggestion(s) (Sniezek & Buckley, 1995; Sniezek & Van Swol, 2001). In this regard, one can note the following things: the first one is that these matching measures are insusceptible to variations between the judge's initial and final decision if the post-advice decision does not match the given recommendation. Concerning this matter, a judge that is exposed to a decision task with three answer choices could choose *'A'* as the pre-advice decision. Imagine the advisor's recommendation is *'B'* and still, the judge decides to choose *'C'*, this as well as the situation in which the judge remains with the initial answer of *'A'* is considered as advice discounting. The second observation relates to the phenomena of *hold* and *shift*, related to the cases that the judge's ultimate decision conforms to the advice received. The only difference is expressed in the judge's initial opinion whether it already has been the same as the advice or not. *Hold* is a matching measure where the judge's initial opinion, the advice and the judge's final opinion all represent the same alternative. More specifically, the judge chooses an option, the advisor recommends the exact same option and hence, the judge won't switch to another option but will remain with this alternative. *Shift*, however, characterizes the situation in which the judge chooses an initial option that does not conform to the advisor's recommendation and the decision maker eventually takes the advice. It might appear that the only indicator for a hold measure or a shift measure is whether the judge adopts the recommendation for the final decision and that advice taking only occurs in shifts but it is also the case in holds. For example, the recommendation received might have encouraged the judge in his or her perception that the initial decision was likely to be accurate even though it was made under a level of uncertainty (Bonaccio & Dalal, 2006). This should be taken into consideration as well.

***Judgment Tasks***

To measure advice utilization that comes along with judgement tasks, a formula-based approach as well as a regression-based approach have been established which will be explained in the next section.

*a) The formula-based approach*

Several formulae have been used to determine the advice taking behavior in judgement decision tasks. Although this formula might seem to be the most suitable one, it can only be applied in cases where judges have to state an initial estimate before receiving the advice. Still, it is able to measure the impact of the recommendation on the decision maker's post-advice estimate. The most useful results are collected when the judge's post-advice decision lies within the range between the initial decision and the advice. One explanation was constituted by Harvey and Fischer (1997; see also Sniezek et al., 2004) denoting advice taking as the relation between the circumstances of the judge's post-advice vs. pre-advice decision and the advisor's suggestion vs. the judge's pre-advice estimate:

$$Advice\ Taking = \frac{judge's\ final\ estimate - judge's\ initial\ estimate}{advisor's\ recommendation - judge's\ initial\ estimate}$$

In addition to that, Yaniv (2004a) stated a similar formula to measure advice taking, naming it the weight of advice (WOA). The only difference between the formula by Harvey and Fischer (1997) and this WOA formula is that the latter only uses absolute values. In a mathematical matter, the formula is displayed as follows:

$$WOA = \frac{|judge's\ final\ estimate - judge's\ initial\ estimate|}{|advisor's\ recommendation - judge's\ initial\ estimate|}$$

A similar method that has been applied by Yaniv and Kleinberger (2000) is the so-called weight of own estimate (WOE) which also includes only absolute values. In this case, advice taking represents the relation between the circumstances of the advisor's recommendation vs. the judge's final estimate and the advisor's recommendation vs. the judge's initial estimate:

$$WOE = \frac{|advisor's\ recommendation - judge's\ final\ estimate|}{|advisor's\ recommendation - judge's\ initial\ estimate|}$$

First, it has to be noted for all three formulae (advice taking, WOA and WOE) that if the advisor's recommendation and the judge's initial decision have the same value such that the divisor equals zero, no outcome can be defined. This does not necessarily result in the fact that no advice has been taken but still it is difficult to quantify the extent of advice taking. Second, Harvey and Fischer (1997) observed that the value of zero for advice taking can be achieved if the initial decision as well as the final decision of the judge are equal. This would imply that no weight on advice was given or, respectively, no advice was taken by the judge. However if the divisor was of a very small value although not zero, the recommendation might have informed the judge about his initial opinion being almost correct and thus, profitable to hold. Furthermore, zero weight on the own estimate is achieved if the judge's final decision and the recommendation are the same. According to Bonaccio and Dalal (2006) a crucial point in the calculation and interpretation of these formulae is to consider general concerns regarding the outcomes when the advice is being *approached*, *overshoot* or *moved away* from. In the following example, the judge's initial decision is 80 and the advisor's recommendation is 100. In both cases, if the judge's final decision is 95 (and thus, the advice is approached) as well as when it is 105 (and thus, the value overshoots the advice), the WOE reaches a value of 0.25. If the judge's final decision is 95 (thus, the advice is approached) as well as when it is 65 (thus, the advice is moved away from), the WOA reaches a value of 0.75. According to that, these formulae can succumb to vague or undefined values, especially if the final decision of the judge does not lie between the stated values of the initial opinion and the recommendation as it is the case when overshooting or moving away from the advice occur. Nonetheless, up to 95% of the observed data in decision tasks appear to align with the presented formulae as most of the stated judge's final decisions lie within the range of the initial decisions and the received advice (Bonaccio & Dalal, 2006).

*b) The regression-based approach*

This method is based on Brunswik's lens model (1955, 1956). By using the regression-based approach, the judge's final decisions are regressed on all sources of advice (Brehmer & Hagafors, 1986; Harvey et al., 2000; Hedlund et al., 1998; Hollenbeck et al., 1995) and the initial decision is suitable to serve as an independent variable. Using this method grants the benefit of a more accurate analysis on the relation between the judge's initial decision, the received advice and the resulting final decision by the judge. However, it has to act with caution as the measurement of advice taking or, respectively, advice utilization for every decision maker is possible only when each individual is involved in multiple decisions and thus, can provide enough data to conduct a proper examination. On the other hand, the formulae suit the purpose of determining the advice taking behavior of judges when they make only one decision and in turn, in case of multiple decisions made by judges, the formulae can be applied to analyze advice taking for each single decision within each decision maker. Yet, correlations can arise due to obstacles like independency of the advisors if they all are provided with equal informative material or if they have the opportunity for a discussion before giving advice (Bonaccio & Dalal, 2006).

### 2.1.5. Other Components affecting Advice Taking

Many researchers have found that the individual behavior of advice taking or advice utilization is affected by several different elements which can be categorized in different groups. The first group comprises factors that have an impact on the decision task itself. These are the task difficulty, the study structure and as well the participant's perception about cost and benefit of taking the advice, i.e. if there is the option for an initial decision before obtaining the advice (Gino & Moore, 2007; Godek & Murray, 2008; Rader et al., 2015; Yaniv & Choshen-Hillel, 2012). Another group comprises the personal factors of the advisor, i.e. characteristics like the advisor's expertise, the confidence level and how the advisor exerts it (Meshi et al., 2012; Sniezek et al., 2004; Tzioti et al., 2014). The third group includes the personal factors of the judge, i.e. characteristics like the knowledge and confidence level (Gino & Schweitzer, 2008; See et al., 2011; Tost et al., 2012). The last group contains elements such as the mutual trust between the decision maker and the advisor or the existence of rewards for either side (Gino &

Schweitzer, 2008; Yaniv et al., 2011). Referring to the last group, trust represents an essential component in advice utilization when it comes to decisions on the basis of value judgments rather than precise correct answers. When facing judgmental tasks, decision makers fix trust in the advisor on similar values (Van Swol, 2011). In tasks that require certain correct answers judges prefer the advisor's expertise and confidence in order to decide whether or not to take the advice (Sniezek & Van Swol, 2001; Van Swol & Sniezek, 2005). Also other researchers confirmed the significance of trust in the advisor in the subject matter of advice utilization (Maister et al., 2001). They showed that the creation of trust in professional advisors is created by individuals exactly like it would be the case in any other relationship. This involves listening, engaging, caring and reciprocity. With regard to incentives, Sniezek, Schrah & Dalal (2004) conducted studies to test the influence of reward on the advice taking behavior by declaring the participating judge as responsible for the reward distribution among two groups where one group received the reward pre-advice and the other group post-advice. The results showed that when including rewards and expertise in the experiment, high levels of advice utilization and post-advice accuracy occur. When investigating also the timing of reward distribution, results revealed that a pre-advice monetary reward distribution will lead to a higher utilization of knowledgeable advice compared to post-advice reward distribution.

The subsequently discussed studies encompass the most frequent contents in the field of advice utilization regarding authority (or power), confidence, inexperienced and experienced advice. It is useful to have a closer look on the findings to get an insight into the case how and to what extent the advice taking process is influenced. The strong impact of power and confidence on advice taking has been highly emphasized by See et al. (2011). In four studies they revealed the importance of these two factors in the matter of judgment accuracy and subject's advice taking behavior. In their first study, subjects with a high self-perceived power level within their job position were assessed by their peers as not willing to take advice. This relationship was mediated by the confidence variable. The second study revealed a positive correlation between self-perceived power and the confidence level in judgments which, in turn, indicated a lower advice taking behavior that can be traced back to a certain unwillingness of participants to change their initial choice. An interesting procedure was chosen for the third study as

some of the subjects were primed to have a higher power self-perception while other subjects were primed to have a lower power self-perception. The first group exhibited a higher confidence level regarding their initial decisions while the other barely indicated any tendency towards or against advice taking. Participants in the fourth study gave less accurate choices when displaying higher levels of both confidence and power than did the control group. The consequential results of these studies indicated that higher levels of self-perceived power and confidence lead participants to overweight their own knowledge and underweight the advice they had been given. This phenomenon is likely to have significant impact on the decision maker's judgment accuracy as the willingness to change one's own initial decision decreases. Tost, Gino and Larrick (2012) also came to the conclusion that power is negatively correlated with advice taking. In addition, they found that participants exhibiting lower self-perceived power levels put greater weight on advice from experts than from unexperienced advisors.. Still, unexperienced as well as expert advice was rejected by participants having a strong self-perception of power and optimism regarding their initial decisions resulting in overconfidence and thus, unwillingness to take advice, especially when they felt to compete with an expert counterpart.

Paying for advice is not uncommon for institutions or individuals when facing decision tasks that are perceived to be unbearable without consulting experts or knowledgeable authorities. Godek and Murray (2008) did research on the behavior of these mentioned players when deciding whether to hire a professional consultant and if so, how much money are they willing to pay for this service. The so-called *willingness to pay* phenomenon thus argues that the decision for or against a consultation is made pre-advice or respectively before the advisor is providing the service. Furthermore they tested the relationship between the willingness to pay and information processing by comparing the experimental and rational type of information processing. While experimentally processed information is based on faster and spontaneous decision making, rationally processed information is slower but based on a logical and justified thinking processes. The study displayed that participants who possess a rational thinking structure were unhesitating to pay for advice compared to subjects possessing an experimental thinking structure. It is noteworthy that the participant's knowledge about the decision issue served as a moderator. This resulted in differences while comparing the two groups when the subjects had only low

specific knowledge available. Further research on the cognitive processes of participants during advice taking has been made by Meshi et al. (2012). The study where participants were told to estimate prices for New Yorker apartments revealed that advice given by experts was more valuable for the participants than advice given by advisors with less experience. Also advice taking was higher when the difference between the judge's initial estimation and the advice was small. Possible explanations for these findings are that judges expect to achieve a better outcome when accepting the expert advice compared to a novice advice and that a small distance in opinions suggests an accordance between the judge and the advisor. The researchers also examined the time that the participants needed to make their final decision. Results were that the smaller the opinion difference was, the longer was the reaction time of the judge. This might be due to the fact that the judge is confronted with a more difficult decision due to that similarity and thus, a more exact decision is pursued.

According to these findings, the researchers found advice taking to consist of several steps which are the evaluation of the advice, the assessment of the differences or similarities between the decision maker's opinion and the advisor's one, and the evaluation of these two steps leading to advice utilization. A similar research area in the literature of advice taking is the justification of advice. Tzioti et al. (2014) examined the participants' reaction to advice based on intuitive or analytical thinking. It was interesting to see that judges discounted the advice when it was based on intuition even if the advisor's expertise level was the same as the one of the judge. In addition, the control group valued advice more even if it was not justified compared to advice based on intuition. On the other hand, in cases where advice that was given by a person with a higher level of experience than the judge, no significant differences regarding the advice being based on either intuitive or analytical thinking were given. These results indicate a relationship between the level of the advisor's expertise and intuitively justified advice but not with analytically justified advice.

Not only do the emotions of the agents play an important role regarding the decision process but also characteristic intersections. According to Yaniv et al. (2011) to use those intersections in a beneficial way, individuals can form self-predictions by considering other's opinions, learn from their experiences

and apply them to the own outcome. The authors furthermore define other's opinions as either the opinion of the majority or of similar others. The impact of the opinion of the majority depends on the size of the group and is affected by social pressure and the belief that the average opinion is trustworthy. In addition to that, similarity is defined in intersections between two or more individuals regarding i.e. their gender, age, religion or ethnicity as well as their status in the society like the type or level of education. In connection with similarity the authors also did research on taste discrimination by the judges and found that decision makers did value advice more if they came from advisors with a similar behavior then themselves while, in turn, participants showing a low score of discrimination still considered advice from a general majority even though at a lower level. By means of their studies the authors concluded that judges indeed have a preference towards advice given by individuals with similar demographic characteristics, still, advice was taken from both similar advisors and the general majority.

Referring to another crucial factor with a high influential potential, Gino and Moore (2007) observed the moderating effect of task difficulty on the advice taking behavior and the opportunity to receive advice by conducting two experiments. They hypothesized that participants choose different strategies in their decision process about the amount of advice to accept. This decision might vary with the difficulty level of the task. Thus, in both studies the subjects were confronted with estimation tasks. In the first study the subjects were asked to state their initial estimation about the weight of the person that was shown in the picture. Next, they had to specify their level of confidence by stating an upper and lower bound interval of which the subjects were 90% sure that the correct answer lies in-between. This was followed by the advice given on the correct answer for the task and the opportunity for the judge to adjust the initial decision and to give a final one. The task difficulty was created by blurring some pictures. Like other studies in this field of research the outcome of the first study by Gino and Moore (2007) revealed that participants tend to weight the received advice significantly higher when solving difficult tasks compared to easy tasks. The second study had a similar experimental design but this time, participants had the opportunity to choose whether or not to receive advice. It was hypothesized that advice seeking would occur more in difficult tasks than in easy tasks, however the findings displayed no significant difference between the complexity conditions. This can be justified by a so-called *ceiling*

*effect* which is due to the advice seeking proportion in difficult tasks and in easy tasks. According to that, 97% of the advice was sought in easy tasks. Interestingly, considering the data provided about the participants' level of confidence, the amount of advice sought was higher than expected as difficult tasks led them underweight their own performance and thus, to take more advice.

### 2.2. Judgment Accuracy

The judgement accuracy represents the accuracy of the output or final decision of a judge. It is also linked to advice taking, implying that the higher the advice using behavior the higher the decision accuracy (Gardner & Berry, 1995; Sniezek et al., 2004; Yaniv, 2004b). Brehmer and Hagafors (1986) argue that the reason for that might be the reduced complexity as judges rely on recommendations in the decision task whereas the advice given does not even have to be perfectly accurate. Also, considering numerous advice of uncorrelated advisors can increase the judgement accuracy due to a reduction in random errors related to the single recommendations (Yaniv, 2004a, 2004b) as forecast combination results in decreased variability and approaches the 'true' value.

The post-advice judgement accuracy of a decision maker can be influenced by three JAS-level variables which are: the average amount of task information that advisors can access, the average advisor's accuracy in their suggestions, and the weight allocated by the judges to the suggestions (Brehmer & Hagafors, 1986; Hedlund et al., 1998; Hollenbeck et al., 1995, 1998; Humphrey et al., 2002). The accuracy of an advisor's recommendation increases with the amount of access to important informational data about the task. This, in turn, influences the judge's accuracy as the decision maker is then able to distinguish between good and bad recommendations and thus, allocates more weight to the first one. Fischer and Harvey (1999) found that a decision maker's accuracy can be enhanced by giving feedback on it or, more specifically, feedback on post-advice accuracy helps the judge to refine the weighting strategy (Hollenbeck et al., 1998). Further research implies the superior effect of feedback on tasks where judges have access to related and easy understandable information about their own performance, for example advice-taking tasks (Harvey & Fischer, 2005).

The level of intelligence and the conscientiousness have been found to be other influencing factors on the decision accuracy. The higher these two variables for both the advisor and the judge the higher the decision accuracy of the latter (LePine et al., 1997). Several researchers found evidence in advice taking leading to better judgment accuracy (Gino & Schweitzer, 2008; Minson et al., 2011; Sniezek et al., 2004) but still, the sense of utilization of the received advice seems to be imperfect (Harvey & Fischer, 1997; Yaniv & Kleinberger, 2000) as judges could have obtained better results than they actually did (Minson & Mueller, 2012).

Another study conducted by Yaniv and Choshen-Hillel (2012) pays attention to the research on decision improvement under circumstances of suspending judgment. The authors argue that judges don't consider others' opinions as they are too influenced by their own. Thus, eliminating the latter would lead to increased judgment performance. To test for this effect, participants encountered estimation tasks where they had to estimate the caloric value of food and were then tested for judgment accuracy and level of confidence. For all three studies the participants were divided into two groups. Considering study one, the first group experienced the full-view condition in which they were allowed to indicate estimates for the food of which they saw the full name before seeing a list with multiple advisory estimates randomly drawn from a survey conducted before. Afterwards they had the opportunity to adjust their initial opinion and state their final one followed by the indication of their confidence level. Participants from the second group experienced the blindfold condition and could not state their estimates before receiving a list with multiple advisory estimates randomly drawn from a prior survey. Furthermore, they were not provided with the full name of the food but rather a letter code and were told that the advisory estimates were stated by participants who knew the full food names. In every round after receiving advice, the participants had to indicate their estimates and the level of confidence. The results showed that the judgments made by the second group under a blindfold condition were more accurate compared to the full-view participants. Hence, there was a positive influence of non-existing opinions prior to the advice as participants without a pre-advice opinion performed better regarding the accuracy of final estimates. The authors based this outcome on the blindfolded subjects exhibiting an unbiased weighting of their

peers' opinions and thus choosing an equal weighting. On the contrary, the full-view participants stuck to their initial opinions and did not benefit from their peers' information. The level of confidence in estimates in the blindfold group was lower than in the full-view group for all three studies conducted. In the second study by Yaniv and Choshen-Hillel (2012) none of the participants, neither in the full-view condition nor in the blindfold condition, were allowed to state pre-advice estimates but rather had to state their opinions after receiving advice and state their level of confidence. Here as well, blindfold participants achieved more accurate judgments by using the judgmental weighting strategy which is similar to the equal weighting strategy. In fact, statistical equal weighting of judgments has been shown to result in more accurate forecasts due to their aggregation compared to individual opinions (Larrick & Soll, 2006).

Other studies showed that decision makers who form initial opinions before receiving advice are able to improve judgment accuracy when seeking for additional information but still receive disadvantageous results due to their inability to benefit from the *wisdom of others* which describes the information or knowledge enclosed in others' opinions (Harvey & Harries, 2004; Mannes, 2009; Yaniv & Choshen-Hillel, 2012; see also Yaniv & Milyavsky, 2007). This effect is enhanced when judges face multiple other individuals' opinions and ascribe the same weight to all of them (Yaniv & Milyavsky, 2007).

Yaniv (2004a) published another paper investigating the use of advice, the weighting of own opinions and the influence on judgment accuracy. Here, he states that a crucial motivational factor for even seeking for advice is the judge's expectation of enhancing the judgmental accuracy and that it is possible to improve forecasts by considering all received estimates (Armstrong, 2001; Ashton & Ashton, 1985; Libby & Blashfield, 1978; Sniezek & Buckley, 1995; Sniezek & Henry, 1989; Sorkin et al., 2001; Winkler & Poses, 1993; Yaniv, 1997; Yaniv & Hogarth, 1993; Zarnowitz, 1984). He mentions that the measurement for accuracy is the mean absolute error of the judgment-criterion correlation (Yaniv, 2004a) and further notes how advice utilization can affect the improvement of quantitative judgments. The studies conducted in the paper were similar to the ones from the paper of Yaniv and Kleinberger (2000) and differed to the extent that participants were classified in knowledge groups. The results

confirmed once more that accuracy is increased when having the option to receive advice. While the participants classified to the low-knowledge groups showed a higher beneficial factor regarding advice, still, the relationship between the knowledge groups and the error type showed no significance (Yaniv, 2004a).

### 2.3. Confidence & Overconfidence

The confidence demonstrated by the involved participants plays a key role in the JAS. Here, one can regard the judge's pre-advice confidence and the advisor's confidence when giving the advice as the input and the judge's post-advice confidence as the resulting outcome (Bonaccio & Dalal, 2006). Confidence can be declared as the forecast of the accuracy of a decision, opinion or suggestion or also as the subjective confidence interval which is the expected value range comprising the correct value. Fleming and Daw (2017) define confidence as the *'(subjectively estimated) likelihood of being correct given the evidence and the decision made'*. This can be measured through indications of estimations or a rating on a Likert-type scale (Klayman et al., 1999). Related to that, the confidence expressed by the advice giver on a decision task can provide the judge with important information about the knowledge, expertise or accuracy of the advice giver (Sniezek & Buckley, 1995; Sniezek & Van Swol, 2001). In this case advice from more confident advisors is less likely to be discounted compared to less confident advisors (Phillips, 1999; Sniezek & Buckley, 1995; Sniezek & Van Swol, 2001; Van Swol & Sniezek, 2005; Yaniv, 1997). Interestingly, overconfident advice givers seem to be favored by judges (Price & Stone, 2004) which allows advisors to adjust their expressed confidence levels in a strategic manner in order to influence the decision makers' judgment (Hollenbeck et al., 1995; Sniezek & Buckley, 1995; Yates et al., 1996). Still, here too are some boundaries to the advisor's ability of boosting the confidence level as advice being too accurate may be considered as impulsiveness instead of confidence (Yates et al., 1996). In any case, Sniezek & Buckley (1995) argue that agreements between the judge and the advisor are more effective in the decision-making process than confidence levels.

Cooper (1991) put a focus on the pre-advice and post-advice confidence of judges exhibited in decision tasks and found that their pre-advice confidence level is dependent on the number of advisors that are sought for advice, leading to the fact that the lower the judge's confidence level the more advice is sought. Regarding the judge's post-advice confidence it can vary due to different variables. On the one hand, judges show an increased level of post-advice confidence the higher the advisor's accuracy (Budescu et al., 2003), on the other hand their confidence levels can rise as they seek advice from multiple individuals or when the advisor's recommendation is based on expanded information. A negative influence appears when the advisor shows conflicts in the given advice so that the judge's post-advice confidence level decreases (Budescu et al., 2003; Budescu & Rantilla, 2000; Sniezek & Buckley, 1995). This effect is intensified when the judge assumes that the advice is based on the same information (Budescu & Rantilla, 2000). Interestingly, the interaction process between judge and one or more advisor(s) causes the decision maker to create a justification in the decision task (Heath & Gonzalez, 1995; Savadori et al., 2001), resulting in a higher level of the judge's post-advice confidence compared to the level of pre-advice confidence (Heath & Gonzalez, 1995).

Overconfidence, however, is defined as a decision maker's excessive self-confidence. This self-confidence can be related to one's own capabilities or achievements as well as to the own knowledge. Being excessively self-confident means to belief that one's own decisions are more precise than they actually are (Gillenkirch, 2018). Another definition arises from Conger and Wolstein (2004) who state that overconfidence describes the individual's overestimation of own skills and the likelihood of successful outcomes as well as the knowledge accuracy. They further argue that these phenomena evolve from the individual's ignorance of the imperfection of their skills. When examining more closely the literature in the field of overconfidence, one can see that the majority of studies and research subjects concentrate on skills at the cognitive level such as forecasts of evaluation tasks according to social characteristics. There is a wide range of research in the field of cognitive psychology that pioneered the construct of overconfidence. Most robust statements in this area argue that individuals are likely to overestimate their possessed skills, their expertise and the accuracy of information they have. According to a study conducted by Bar-Hillel (2001), participants that are a certain percentage sure about their

response being correct, they actually are correct in less than that certain percentage of the time, on average. There is a plenitude of reasons for individuals to exhibit overconfident behavior and indeed various investigations have revealed the occurrence of overconfidence in many structures such as students (Zakay & Glicksohn, 1992), entrepreneurs (Baron, 2000) and negotiators (Neale & Bazerman, 1991), just to name a few.

Taking a closer look on the issue of overconfidence, there also has been done plenty of research on judges showing overconfidence in decision tasks (Klayman et al., 1999), but only few researchers have focused on this phenomenon in the Judge-Advisor system, amongst which are for example Sniezek and Buckley (1995). They observed that judges would show overconfidence in their judgements when having the fewest amount of information about the given task and thus were forced to trust in the advisors' recommendations. Although this had no effect on the judgement accuracy, still higher levels of the judges' confidence could be observed when the decision makers were interacting with each other and had the opportunity to exchange received recommendations (Heath & Gonzalez, 1995). A crucial point here is that the occurrence of overconfidence is highly dependent on the type of task that the judge is confronted with. Klayman et al. (1999) as well as Soll and Klayman (2004) found that the mentioned phenomenon is more likely to occur in judgement tasks compared to decision tasks.

Judges that show signs of overconfident behavior regarding their personal knowledge and expertise tend to underweight the necessity of advice. As overconfidence describes the perception of one's own knowledge higher than it actually is or the exhibition of a higher confidence in own skills (Sniezek & Buckley, 1995), overconfident people often believe that their decisions are more accurate than those of others (Larrick et al., 2007; Moore & Healy, 2008) and thus, are less addressable to seeking and using advice (Gino & Moore, 2007; Harvey & Fischer, 1997; Soll & Larrick, 2009). In addition, literature found elements like power, perceived competence and narcissism to be positively related with overconfidence (Campbell et al., 2004; Fast et al., 2012; Larrick et al., 2007; Moore & Schatz, 2017; Schaefer et al., 2004). Other elements such as monetary stimuli or feedback at the right time are able to

decrease overconfidence. This is due to the scarcity of economic prerequisites and the double meaning of traits (Proeger & Meub, 2014).

Moore and Healy (2008, p. 502) argue that there are *'inconsistent ways'* to characterize overconfidence and pioneered a more specified distinction regarding the the three types overestimation, overprecision and overplacement. But these types are not to be seen as fixed and standardized. They investigated existing issues in the subject matter of overconfidence and stated that the emergence of problems is due to the lack of differentiation between various types of overconfidence. The first issue that the authors mention in their analysis refers to the fact that substantial research often mistake overestimation and overprecision. The second problem addresses the evolving underconfidence that describes the phenomenon in which individuals underestimate their own knowledge and performance and place themselves lower compared to others. The last problem stated by the authors is the occurring inconsistency between overestimation and overplacement. Here, the achieved findings point out a negative relation between the two types overestimation and overprecision with regard to the task difficulty. They argue that easy tasks are subject to both underestimation and overplacement, and difficult tasks succumb to overestimation as well as underplacement (Moore & Small, 2007). Their analysis further reveals that individuals not only lack sufficient knowledge about their own personal skills and performance but also about those of others which creates expectations that are not correct (Moore & Healy, 2008).

The present paper takes overconfidence into account. The purpose of this approach is the paper's determination to build on already existing literature in the field of over-/confidence and to see whether task difficulty and performance feedback have an impact also on the three types of overconfidence in the JAS. As the focus of this paper is to examine the effect of performance feedback on the mentioned dependent variables (use of advice, judgment accuracy and overconfidence) on the two levels of task difficulty, the study includes only one new variable for examination, which is performance feedback. As already mentioned in the overconfidence literature before, people that are overconfident are considered to be less willing to accept help and overconfidence is more likely to occur in judgment tasks

(Klayman et al., 1999; Soll & Klayman, 2004). This will be part of the examination in this paper. In the next sections, the three types of overconfidence will be presented and discussed.

### 2.3.1. Overestimation

Overestimation can be observed in tasks where individuals overestimate their own performance (Moore & Schatz, 2017), for example when individuals believe to have a certain amount of correct answers in a task although the actual amount of correct guesses is below that belief (Moore & Healy, 2008). In a similar study of Moore and Dev (2018) participants were confronted with tasks about general knowledge and had to indicate how many answers they got correct as well as their level of confidence which allowed the researchers to test for overestimation. Overestimation does not only occur in estimation tasks but also when it comes to predictions about certain events. Magnus and Peresetsky (2018) conducted a study with statistics students and examined their level of overestimation in three exams. After the first part of each exam the students had to state forecasts about their final grade in that very exam they just had finished. As a motivation, students with the most accurate forecast were granted a bonus in their grades. Findings showed that participants with a comparatively bad forecast performance presented a higher level of overestimation regarding their grades compared to students with a better forecast performance. An interesting circumstance of the experiment that this study provides is that the participants had to undertake more than one exam which allowed for a more accurate testing of the development of the level of overestimation as well as whether the students were capable to improve their judgment. Regarding the students' level of overconfidence, it was found that this variable decreased over time as the participants aligned their forecasts. Also, differences in genders were found where females showed a more rational behavior with a lower level of overconfidence as well as revealed a faster learning process in their forecast decisions. According to that, the authors state that most students were susceptible to overestimation but both males and females were able to adjust their forecast decisions.

Other influencing factors in the matter of overestimation are mentioned by the literature to be feedback and rewards. In a study by Bolger and Önkal-Atay (2004) students had to state interval forecasts for

stocks and were provided with feedback on their performance. It was found that feedback can decrease overestimation. One explanation for the existence of overestimation is the fact that participants fail to assess properly their level of control in terms of future outcomes (Gino et al., 2011; Moore & Schatz, 2017). Also, overestimation is seen when participants rate their speed of task accomplishment much higher than it really is, resulting in the *planning fallacy* effect (Moore & Schatz, 2017).

### 2.3.2. Overprecision

The phenomenon of overprecision describes the event in which participants overweight their own performance, namely when they are too confident in the correctness of the answers given (Moore & Schatz, 2017). According to Moore, Carter and Yang (Moore et al., 2015) the appearance of overprecision can be due to anchoring, conversational norms and the naïve intuitive statistics. The first explanation describes the situation in which participants, based on their first guess, fail to adjust their second guess sufficiently while the second explanation indicates that participants lack accuracy due to their aim for informative judgments. The third explanation states that actual error distributions are higher than subjective ones as only a small part of important facts is cognitively absorbed. Other researchers see the cause for overprecision in the fact that participants feel challenged regarding their knowledge when they consider themselves as experts in the topic covered by the study questions (Cesarini et al., 2006). To test for overprecision, in most cases participants encounter questions of a quantitative type and have to state at which boundaries they are 90% certain that the right answer falls between these values (Mannes & Moore, 2013; Moore et al., 2015; Moore & Dev, 2018; Moore & Healy, 2008). This indicates the 90% confidence interval and yields more reliable results than the measurement of point estimates (Moore & Schatz, 2017). Still, some disadvantages lie in the method of the 90% confidence interval. First, most participants might not be used to indicate their confidence interval (Mannes & Moore, 2013; Moore & Schatz, 2017). Second, the choice of the questions can influence the research outcomes (Soll & Klayman, 2004). In this regard, to test for overprecision the utilization of two-choice questions in the study is useful as the participants have to decide between the two presented options and afterwards indicate on a scale from 50% to 100% how confident they feel that the answer chosen is right

(Soll & Klayman, 2004). It is noteworthy that in studies testing for the 90% confidence interval the cases in which the boundaries actually contained the correct answer was below 50%. This is due to the fact that participants already supposed their boundaries to be too narrow and, interestingly, that the boundaries chosen by other participants were incorrect. In turn, when participants are provided with the option to adjust their confidence intervals, usually the boundaries were expanded to be sure that the correct answer lies in-between (Cesarini et al., 2006; Klayman et al., 1999; Moore & Healy, 2008; Soll & Klayman, 2004).

### 2.3.3. Overplacement

Overplacement can be defined as a strong belief about oneself being better than others or amongst the best performers which is also called the *better-than-average* effect (Moore & Dev, 2018; Moore & Healy, 2008; Moore & Schatz, 2017). Focusing on it, the testing measurement for overplacement is often applied via a scale rating whose answer options for the participants' performance range from 'much worse than average' to 'much better than average' or also from 'below the worse 10%' to 'above the best 10%'. Other researchers use specific quantifiable measures to test the participants' sense of comparison to others (Moore & Schatz, 2017). Zell and Alicke (2011) as well established a similar study where participants had to undergo a self-evaluation test and define their traits and capabilities amongst others on a rating scale compared to an average participant. According to some researchers, the *better-than-average* effect is likely to occur in comparison situations between participants and an average person as participants are prone to choose the one comparison target of which they think has a poor performance in the corresponding judgment dimension (Perloff & Fetzer, 1986). Another explanation addresses the subjects' selective thinking process regarding their own performance (Weinstein, 1980). Contrary to these statements, Alicke et al. (2001) emphasize that comparisons in behavior do not necessarily justify the effect. The proper measurement of the overplacement phenomenon is affected by several elements which are the characteristics of the judge and the comparison target, the rating scale, and the judgment dimension (Alicke & Govorun, 2005). According to the literature, the characteristics of the judge and the comparison target, or more precisely, the participants' self-esteem, play an

important role regarding overplacement. It was found that participants indicating a higher level of self-esteem also placed themselves higher in the self-evaluation scale (Brown, 1986, 2012). In addition, the mental state of participants was found to be an influencing factor as well. According to that, participants who experienced a condition of depression tend to evaluate themselves more at an average, yet, these findings are not only applicable on depression relevant attributes like discouragement or frustration but also on depression-irrelevant attributes like honesty and daily life activities (Alicke & Govorun, 2005; Tabachnik et al., 1983).

Testing methods for overplacement can be of a direct or indirect type. With the direct method, a form of the *better-than-average* effect can be observed when the participants' self-evaluation against an average participant is above the middle value of the provided scale. Furthermore, a direct scale allows for a widened comparative framework and is applied in more studies, though it succumbs to a worse explanatory nature compared to indirect scales (Alicke & Govorun, 2005; Zell & Alicke, 2011). Besides, direct comparisons diminish the effect of overplacement (Otten & Van Der Pligt, 1996). In turn, with the indirect method, participants as well as the average person indicate their traits rating on different scales before being compared (Zell & Alicke, 2011). Another crucial factor in the measurement of overplacement is the sequence of questions as the fact whether participants have to evaluate themselves or other individuals first can induce significant differences in the results (Stavrova et al., 2016).

Concerning the mentioned influencing factors, some characteristics can be more objective rather than subjective and vice versa, such as when characteristics comprise either a social or moral, an intellectual or physical nature, or are based on definitions by other participants (Alicke et al., 2001; Alicke & Govorun, 2005). A remarkable result of a study by Alicke and Govorun (2005) is that participants saw themselves as superior in controllable attributes and thus, had higher ratings in positive traits while they saw themselves inferior in uncontrollable attributes, indicating lower rating in negative traits. This supported on the one hand the findings of Brown (2012) that the *better-than-average* effect is most likely to occur in rating tasks when it comes to more important characteristics such as honesty and

kindness, and on the other hand the results of Pedregon et al. (2012) who found that the effect is stronger when rating attributes that are perceived as socially desirable.

For purposes of a holistic understanding, at this point also the *better-than-myself* effect is mentioned which was discovered when participants took part in a study where they were confronted with a self-evaluation task regarding trait dimensions. More specifically, participants had to evaluate the percentage of how many times they showed a cooperative or uncooperative behavior when possible. About six weeks later, the study subjects were provided with results of which they believed they portrayed the average behavior percentage. In fact, these were the same estimates the participants gave in the first round. For example, when a student indicated a value of 86% of cooperative behavior and a value of 14% of uncooperative behavior, he or she later believed that the average participant was cooperative in 86% and uncooperative in 14% of the cases. Based on these respective numbers, participants were asked once again to estimate their own performance compared to the average. The study showed that the second self-evaluation was higher than the communicated average result although those were the students' own estimations (Alicke & Govorun, 2005). As additional information it is noteworthy that the counterpart of overplacement is called underplacement and also describes the *worse-than-average* effect. It occurs when participants underweight themselves compared to the average participant (Alicke & Govorun, 2005; Moore & Healy, 2008; Moore & Schatz, 2017).

### 2.4. Task Difficulty

Gino and Moore (2007) were interested in the influence of task difficulty in estimation tasks. They conducted two studies where participants were shown pictures of people, had to estimate their weight and received recommendation on the correct answer before having the chance to adjust their decision. The given advice represented the best guesses collected from a randomly chosen participant from previous rounds. It is noteworthy that Gino and Moore included payments in their studies such as a $10 show up fee and the chance to win up to $20 depending on the accuracy of the participants' estimates. The experimental setting included the application of task difficulty. In both studies the difficult condition

was established by blurring some of the pictures to increase the level of difficulty for the subjects. More precisely, every odd round introduced a new picture to the participants while every even round contained the same picture but this time the advice on the correct answer was provided as well. In the next odd round the participants were shown a new picture but in a blurred setting representing the difficult condition. Again, the subsequent even round contained the same blurred picture as in the round before as well as the advice. Study 1 provided the participants with the advice although they did not ask for it while study 2 enabled the subjects to choose whether or not to ask for advice. The background of the first experimental design was that it allowed for a proper testing of whether participants value the received advice more if it was provided and made the subjects believe that the experimenter thinks of the advice as being an important element in the decision process. The second experimental design was suitable for testing whether participants even consider to seek for advice or not. The third testing by the researchers included the difference in advice taking between these two studies, meaning whether or not the utilization of advice is higher in cases where participants decided on their own to ask for it or whether they were obliged to receive it. Before starting with the study, the subjects were allowed to perform four practice rounds to get used to the tasks. After each even round of the original study, the subjects were in addition asked to provide information on their 90% confidence interval.

The obtained results from study 1 where participants received advice by default showed that subjects ascribed a higher willingness to take advice in difficult tasks than in easy tasks although the fact that advice was obligatory provided may have influenced the use of advice as in general participants do have the opportunity to choose whether or not to ask for it. Still, this was explained by the fact that participants did not think of their own information and knowledge being useful to them in the particular task. Further results indicated that subjects did not express equal levels of confidence in the two difficulty conditions and, in fact, that the demand for perceived useful information was higher in the difficult condition compared to the easy condition. Additionally, Gino and Moore (2007) found in a mixed model analysis that advice weighting in difficult tasks is higher than in easy tasks due to the participants' decreased level of confidence in more complex settings. More precisely, advice was underweighted more in easy tasks than in difficult tasks while advice was overweighted more in difficult tasks than in easy tasks.

The second study incorporated the possibility for participants to choose whether or not to receive advice on the correct answer. More precisely, in each even round where they were shown the same picture as in the previous round the subjects were not provided by default with the advice but rather saw a button appearing that would provide advice on the weight when pressed. If participants did not wish to receive advice they continued with stating their estimates without pressing the button. Findings of an analysis of variance showed that the subjects were not equally confident in their estimates regarding the two difficulty conditions and expressed a higher demand for information in difficult tasks than in easy ones. Also, the weight of advice in difficult tasks was higher than expected compared to easy tasks, considering the participants' level of confidence. Although their results were consistent with their hypotheses, namely that participants would seek for advice more in difficult tasks than in easy tasks, the authors did not achieve any considerable differences with regard to the participants' advice seeking behavior. In order to answer the third research question, namely whether advice utilization is higher in cases with advice given by default or optionally, an analysis of variance was conducted, revealing that advice utilization is higher in conditions with difficult tasks than with easy tasks.

Brooks et al. (2015) focused on the subject matter of advice seeking and found task difficulty to be a moderating factor, amongst others. In their research on the moderating roles of task difficulty, advisor ego and advisor expertise they argue that task complexity represents a common moderator of the relationship between advice seeking and the perceived competence. In terms of the focus on task difficulty, only the third study will be discussed. Here, the researchers expected a higher increase in perceived competence due to advice seeking in difficult tasks compared to easy tasks. In four between-subject conditions (easy task vs difficult task and advice-seeking vs no-advice-seeking) the subjects were asked to solve either an easy or a difficult task. While the difficult tasks implied the work on brain teasers, the easy tasks did contain rather simple addition tasks. In both conditions the opportunity to send instant messages to an assigned partner which was in fact a computer-simulated confederate was given. In the end, participants had to rate the confidence of their partner. The results displayed that the score of correct answers was higher in the easy condition than in the difficult condition. Regarding the easy tasks, the evaluation of competence showed no significant differences concerning the two advice-

seeking conditions. In summary, results showed a moderating effect of task difficulty on the relationship of advice seeking and competence perception as a higher task complexity led to a higher competence perception. Nonetheless, advice seeking had no impact on competence perception in easy tasks. This finding has already been supported by Gino and Moore (2007) who argue that participants are most likely to seek for advice when they perceive their own knowledge as insufficient and when they face difficult tasks. According to Brooks et al. (2015) not only is advice seeking perceived to improve the decision accuracy in difficult tasks but also the ones who seek advice see themselves as more competent due to their decision for seeking advice as it represents a self-reflection regarding personal limits which can lead to improved decision making. In turn, easy tasks can mislead the advisor to think that the advice seeker is incompetent which might induce a decreased willingness to seek for advice.

In a different experimental setting, Zhou et al. (2020) examined the effect that time pressure and task difficulty have on human search performance. In their study, thirteen participants were exposed to time pressure when the time that the target information was presented on the screen was varied and they experienced task difficulty when the amount of target information shown on the screen was varied. The experimenter collected data about the participants' reaction time, the accuracy rate and eye movement. Related to the subject matter of the impact of task difficulty, they found that the participants showed a higher reaction time and a lower accuracy when tasks became more difficult. Moreover, pupil diameter and task difficulty were found to be positively correlated and the researchers argue that time pressure and task difficulty presented crucial factors in the field of search performance.

### 2.5. Performance Feedback

Due to the rare research in the field of performance feedback in relation to advice taking, judgment accuracy and confidence altogether, this section covers only the most important publications that contribute to a basic understanding of the topic.

Yaniv and Kleinberger (2000) conducted a research on the subject matter of weighting techniques by decision makers or, more precisely, how judges consider advice and adjust their decisions based on others' advice. The first study included feedback performance and judgment accuracy and was composed of two phases. In the first one, participants had to specify their estimates for each of the presented fifteen questions as well as their 95% confidence interval. The following second phase included the same questions for the same group of participants but this time the judges could choose between the presented own estimates from phase one and a randomly chosen advisor's opinion retrieved from a prior study before stating their final answer. A feedback informing about the correct value was shown instantly after each final answer. This procedure was advantageous in order to allow the participants to review their own judgment accuracy as well as that of the advisor. A reward regarding the accuracy was provided for the participants. In this first study it was found that by considering the opinion of an advisor participants could improve their judgment accuracy which might be due to the accuracy feedback they received. The second study centered the research question whether participants adjust their weighting policies according to good or poor recommendations as well as whether the receipt of online feedback had a significant influence on the adjustment of weighting policies. More precisely, the authors investigated the participants' ability to differentiate between useful and less useful advice and the impact of existing or non-existing feedback on their accuracy. This time again, the online feedback was provided right after the subjects indicated their final estimate. The results showed that the participants' WOE variable was higher in cases with a poor recommendation. Hence, feedback enabled the participants to distinguish better between good and poor advice and strengthened the subjects' ability to assess the quality of both the advice and the advisor.

Other researchers showed that receiving performance feedback in a constant manner can help to enhance the participants' learning goal orientation because subjects succumb to the norm of striving for learning progress (Ghoshal & Bartlett, 2007). In relation to that, research in the field of performance systems revealed that performance feedback can boost the participants' ambition for knowledge and sharing (Mathieu et al., 2008). With respect to the linkage to individual goal orientation, feedback is often used by participants to assess their own performance in relation to their personal goals or performance

expectations for the task. Here, not all subjects do react the same to feedback. Some will show a higher willingness to achieve the set goal, some will change their goal and others, in turn, decide to reject the received feedback or even quit the goal (Annosi et al., 2020).

In the field of management, researchers analyzed the effect of different performance feedback signals in managerial decision-making processes on performance improvement at the program level. According to Min et al. (2020), program performance feedback provides insights into the efficacy of processes and the factor flow like input, output and the outcomes and remark the importance of program managers in the process of performance feedback. Though some prior research has already shown that managers seldom take advantage of provided performance information (Kroll, 2013; Donald P. Moynihan & Lavertu, 2012), there is enough research indicating performance information utilization by managers, especially when the provided information is perceived as useful or trust in own performance is low (Gerrish, 2016; Kroll, 2013; D. P. Moynihan & Pandey, 2010). Hence, Min et al. (2020) underline that performance information has to be used to have a positive impact on performance.

### 2.6. Research Gap

The issue of advice giving and advice taking has received active attention from researchers for more than 20 years (Brehmer & Hagafors, 1986). There has been recognized research on the subject matter of advice including the types of advice (Bonaccio & Dalal, 2006; Dalal & Bonaccio, 2010), advice utilization and advice discounting (for example Bonaccio & Dalal, 2006; Yaniv, 2004a, 2004b) and also particularly under consideration of confidence for instance with regard to pre-advice and post-advice confidence (Cooper, 1991). In the field of judgment accuracy it was found that a higher accuracy is positively correlated with advice taking (Gardner & Berry, 1995; Sniezek et al., 2004; Yaniv, 2004b) and that task difficulty plays a fundamental role in advice taking as the more difficult a task is, the more the advice is overweighted (Gino & Moore, 2007).

Yet, to my best knowledge, engagement on the impact of task difficulty and especially of performance feedback together is scarce. As the present paper includes several factors – task difficulty, performance feedback, advice utilization, judgment accuracy and confidence – it is clear that this covers a highly specific area in the field of advice taking. Primarily, only rare research has been done so far regarding the most essential factor of this thesis which is performance feedback. The importance of this variable is justified in the fact that although there is plenty of studies on advice taking behavior in estimation tasks, there is hardly any research focusing directly on the effect that feedback on the judges' performance has on their advice taking behavior. The impetus of this paper is to have a closer look on the so far neglected role of performance feedback in the process of advice taking and to test the effect this variable has on the participants that take on the role of a judge. The aim is to answer the research question whether participants learn to use the advice they receive in a more optimal way over time when being provided with feedback on their performance. In addition, two different task complexity conditions are implemented. In the present paper, an increase in the judgment accuracy is considered as the indicator for a more optimal utilization of advice. The link to advice utilization is presented in the expectation that judgment accuracy is positively correlated with advice taking (Gardner & Berry, 1995; Sniezek et al., 2004; Yaniv, 2004b). Another interesting factor here is the progress of the judges' confidence throughout the task. The findings resulting from this study will provide a first and basic approach to the topic of performance feedback as an additional influencing factor on advice taking. Based on these outcomes, further and more detailed research can be done with regard to possible biases and heuristics.

## 3. Hypotheses

The findings discussed in the literature review of this paper serve as a basis for the establishment of adequate hypotheses to answer the research question whether participants in the roles of judges learn to use advice in a more optimal way over time when they receive feedback on their performance. As already stated, this paper examines the influence that task difficulty and performance feedback have on advice taking behavior, judgment accuracy and the types of overconfidence of the participants.

According to that, the first hypothesis concerns the advice taking variable (WOA) which describes the weight of advice or more precisely, to what extent participants weight the advice they receive and thus, take it into account. Literature in the field of task difficulty has already revealed that participants' advice taking behavior is higher in difficult tasks than in easy tasks and that advice weighting is more expressed in more complex tasks (Gino & Moore, 2007). Furthermore, researchers found that receiving performance feedback is likely to enhance the ambition for knowledge and sharing (Mathieu et al., 2008) and that the learning goal orientation of participants is positively impacted by performance feedback as participants strive for learning progress (Ghoshal & Bartlett, 2007). Based on that, not only the value of the WOA variable is expected to rise in general but also to do so more in difficult tasks than in easy tasks. This means that participants are expected to be more willing to take advice over time and their advice taking behavior should be stronger in difficult tasks compared to easy tasks.

*Hypothesis 1: The WOA variable will show higher values in difficult tasks than in easy tasks over the rounds.*

According to the findings in the literature, performance feedback can improve the decision maker's judgment accuracy (Fischer & Harvey, 1999; Yaniv & Kleinberger, 2000). Also, higher advice taking has been proven in several studies to have a positive impact on the judgment accuracy (Gino & Schweitzer, 2008; Minson et al., 2011; Sniezek et al., 2004). In this regard, for the second hypothesis participants are expected to show a higher accuracy in their judgments but will do so more in difficult

tasks than in easy tasks. It is expected that judgment accuracy will be higher in difficult tasks due to factors such as higher advice taking and thus, improved adjustment of estimates after receiving performance feedback.

*Hypothesis 2: The judgment accuracy variable will show higher values in difficult tasks than in easy tasks over the rounds.*

The third hypothesis comprises the participants' level of overconfidence in estimation tasks. See et al. (2011) found in their studies that high levels of confidence lead participants to overweight their own knowledge and thus, underweight the received advice which indicates lower advice taking. It is assumed that performance feedback will alter the subjects' self-perception in their level of overconfidence. Other research shows that in difficult tasks participants feel less secure in their judgments. This was also confirmed by Gino and Moore (2007) whose studies showed that a higher task complexity results in higher advice weighting due to the participant's decreased level of confidence. It is expected that participants will experience low levels of the types of overconfidence when working on difficult tasks. Therefore, the third hypothesis will be divided into three parts to test for all three types of overconfidence.

*Hypothesis 3a: The overestimation variable will show lower values in difficult tasks than in easy tasks over the rounds.*

*Hypothesis 3b: The overprecision variable will show lower values in difficult tasks than in easy tasks over the rounds.*

*Hypothesis 3c: The overplacement variable will show lower values in difficult tasks than in easy tasks over the rounds.*

The following part of this paper addresses the approach that was applied in order to test for the aforementioned hypotheses. It comprises the elements of procedure, information on the participation as well as included data measurement description.

## 4. Methodology

The analysis of the survey was conducted with an SPSS program. The collected data was corrected such that 103 out of 183 data sets were useful. Entries that were not meaningful, incomplete or obviously too similar were eliminated. The independent variables were illustrated by '*time'* (which indicated the points in time when the pictures of food were presented) as well as '*level of difficulty'*. The dependent variables were illustrated by the WOA variable, the judgment accuracy and the overconfidence which was divided into three overconfidence types that are overestimation, overprecision and overplacement. A MANOVA analysis was used as it is most suitable when working with more than one variable (the M stands for multivariate). The benefits of this multivariate analysis method are that it allows for the investigation on the relations between the dependent and independent variables as well as for the relationship between the dependent variables. The significance level was set for $p < .05$.

### 4.1. Procedure

Although the methodology is based on the experiments conducted by Gino and Moore (2007) as well as Yaniv and Choshen-Hillel (2012) a few modifications on the present analysis have been made such as the consideration of only the situation in which the advice is given to the subjects by default. Thus, they only had the option to choose whether or not to follow the advice but not whether or not to receive it. The other modification concerns the type of object for the estimation tasks. The survey was created via an online tool on www.umfrageonline.com. The participants did not have to estimate the weight of people shown in the pictures but the amount of kilocalories of dishes. This is based on Yaniv & Choshen-Hillel's study (2012) and is considered to be more suitable as pictures of individuals imply more preparation and also data privacy issues. As the task difficulty plays an important role for the subsequent analysis, a hard and easy effect was implemented. This is shown by the fact that in each odd round the participants were shown a list of the exact ingredients per serving below the picture of the corresponding dish, representing the easy condition. This information could have been used as a guideline by the subjects when stating their initial estimate. The hard condition, correspondingly, did not provide an ingredients list such that the participants did not have that information. This procedure represents the

hard and easy effect which is important for the subsequent analysis of how the performance feedback impacts the participant's behavior in advice taking under consideration of the task difficulty level

Before beginning with the estimation tasks the participants were informed about the upcoming survey being part of the achievement of a master's degree and that the subjects will be anonymously asked to give estimates about the amount of kilocalories per serving of the dishes shown in the corresponding pictures with the addition that some of the pictures come along with an ingredients list and some not. The purpose of the study was only vaguely revealed by letting subjects know that the research is about the decision process in estimation tasks. No further information on the investigation on the impact of task difficulty and performance feedback was given in order to avoid the participants being too influenced by the mentioned variables so that they would focus on their estimations only. Furthermore it was explained that the survey consisted of twelve rounds with a lead time of approximately ten minutes, that the subjects would be presented with an estimation of another participant who took part in this survey earlier and that they then would have the opportunity to adjust their initial opinion. This other person's estimates served as the advice and were collected by presenting the study to an independent and randomly chosen person before officially publishing the survey. In this case only the estimates and no answers regarding the overconfidence types were relevant, hence this person was provided with a shortened questionnaire. After providing data about the participant's demographics like gender, range of age, profession and degree of education once more the instruction text was displayed to the participants before starting with the first round. The procedure in each round was the same: the participants were shown a picture of a dish with the original title above. The pictures were retrieved from www.eatsmarter.de, a publicly available website providing recipes. The odd rounds offered an exact ingredients list and thus, represented the easy condition while it was the opposite in even rounds. First, the subjects had to make an initial estimate about the kilocalories for the dish shown in the picture. This was followed by questions to test for the three different types of overconfidence. Overestimation was tested by asking the participants about how many kilocalories the initial estimation might be apart from the correct answer. Overprecision examines the 90% confidence interval and was tested in two steps. It required an upper and lower boundary of the amount of kilocalories such that the participants

were 90% sure that the correct answer lay in-between the chosen boundaries. For a better understanding a sentence in the first-person perspective was stated for each boundary so that there were only two gaps to fill in. This process is used to analyze the subjective confidence in the subjects' judgment (Soll & Klayman, 2004). After that, overplacement was tested by asking the subjects to state their opinion about what they think how close their estimation is to the correct amount of kilocalories compared to other participants in the survey. Here, the task was to range their performance into top, bottom or average percentages (i.e. 'top 10%'). Again, for better understanding a sentence in form of a first-person perspective was given. As announced in the introduction text, the participants were then provided with the advice depicted by the estimates of the random participant that had conducted the study before and the opportunity to choose whether or not to adjust their initial estimation. As a support, the picture and, if applicable, the ingredients list, were provided once more. Depending on their choices, subjects that were willing to revise their choice were redirected to a page to state their adjusted and initial estimate while subject that did not want to take the advice into account were directly led to the subsequent page. Consequently, their initial estimation represented their final estimation. Subjects were then informed about the correct amount of kilocalories per serving of the corresponding dish which served as the performance feedback and were told about the end of the round. After following this procedure for the next eleven rounds the participants were asked two more questions regarding their personal opinion about food kilocalories in their daily life. Here, they had to state whether or not they think in general about kilocalories of the food they eat and if the amount of kilocalories of food plays an important role in their daily life. At the end of the survey the participants had the opportunity to leave comments.

### 4.2. Participants

For testing the aforementioned hypotheses in a proper way at least 70 evaluable surveys were needed. Participants were requested in several ways. Since the survey did not require a certain subject pool, participants were approached via word-of-mouth advertising as well as via the provision of the survey in purposive survey groups on social media platforms or other survey provider like www.surveycircle.com. To gain a higher range for subject recruitment the survey was published in

English as well as in German. As 103 out of 183 people finished the survey until the end, this leads to a dropout rate of approximately 43.72%. Regarding the survey in English language the participation rate was 52.4% (54 subjects) and in the German survey 47.6% (49 subjects) completed the survey. Among the participants of 103 individuals, 35% (36 subjects) were male and 65% (67 subjects) were female. 1.9 % (2 subjects) was under 20 years old, 13.6% (14 subjects) was between 21 and 24 years old, 48.5% (50 subjects) was between 25 and 29 years old, 14.6% (15 subjects) was between 30 and 34 years old, 4.9% (5 subjects) was between 35 and 39 years old, 5.8% (6 subjects) was between 40 and 44 years old, 1.9% (2 subjects) was between 45 and 49 years old, 2.9% (3 subjects) was between 50 and 54 years old, 1.9% (2 subjects) was between 55 and 59 years old and 3.9% (4 subjects) was 60 years or older. Regarding the profession, 42.7% (44 subjects) had a student status, 50.5% (52 subjects) was employed, 1.9% (2 subjects) was self-employed, 1.9% (2 subjects) was retired and 2.9% (3 subjects) was unemployed. With regard to the education, 1% (1 subjects) had less than a high school degree, 16.5% (17 subjects) had a high school degree or equivalent, 38.8% (40 subjects) had a bachelor's degree while 32% (33 subjects) had a master's degree, 5.8% (6 subjects) had a professional degree, 4.9% (5 subjects) a doctorate degree and 1% (1 subject) indicated others. Concerning the questions at the end of the survey about the participants' general attitude towards kilocalories in their daily life 56.3% (58 subjects) did not state to think about kilocalories of the food they eat while 43.7% (45 subjects) did. Also for 71.8% (74 subjects), the amount of kilocalories of food does not play an important role in their daily routine while it does for 28.2% (29 subjects). What was interesting to see is that approximately 20.39% (21 subjects) of the participants that fulfilled the survey until the end left a comment in the provided comment section where most of them were about the participant's surprise about the correct amounts of kilocalories of the dishes.

### 4.3. Data Measurement Description

In order to analyze the impact of task difficulty and performance feedback on the use of advice, judgment accuracy and overconfidence, the variables and their relations had to be measured. As the advice taking or advice utilization variable plays a crucial role in this study, a measurement was used as it has been in

several studies before (Harvey & Fischer, 1997; Yaniv, 2004a): the formula of *weight of advice* (hereafter WOA). This measurement is an indicator for how much the participant uses the received advice and in turn, to what extent the participant discounts the advice (Yaniv, 2004a). The cleared data was subject to the calculation with the described formula for each condition of task difficulty. In this connection, a WOA value of 0 meant that the participant did not weigh the received advice at all but 100% discounted it. The other way round, a WOA value of 1 was achieved when the participant's final estimation was equal to the advice given and thus, the advice discounting was 0%. Also partial discounting was possible when the participant's final estimate took on a value between the own initial estimate and the given advice. Here, the judge weighted both the own estimate as well as the advice positively (Yaniv, 2004a) (i.e. averaging both values would lead to a WOA of 0.5). This resulted in the fact that the WOA variable served as an indicator for the shift rate, meaning the extent to which the judge used the advice.

Furthermore, information about the judgement accuracy was needed to allow for proper insight on the analyses of this study. Therefore, the absolute error of judgment accuracy was calculated. Here, the numerical difference between the correct value of kilocalories of the dish shown and the participant's final estimate was computed, considering the two different task difficulty conditions. The outcomes then showed how near or far the participant's final estimate was from the correct answer where a low error of judgment accuracy represented a high level of judgment accuracy as the participant indicated a final estimate near to the actual correct answer. In turn, high values of the error of judgment accuracy indicated a low level of judgment accuracy as the participant's final estimate drifted away from the correct answer.

To test for the types of overconfidence, four questions were stated as described above in the procedure section. To assess the participants' behavior in terms of overestimation as the first type they had to indicate how many kilocalories they believed their initial estimation might be apart from the actual correct value. Being provided with this information, a comparison between this and the actual correct value of kilocalories was possible. Subsequently, a calculation of the difference between the actual

amount of kilocalories apart and the estimated amount of kilocalories apart was conducted for every round. This indicated either the absence of overestimation if the difference was 0 or the presence of overestimation to a certain extent if the difference took on a positive value. The higher the value the more overestimation was present. In turn, a negative value referred to some sort of underestimation which described the cases in which the participants indicated a higher amount of kilocalories being apart from the correct answer than they actually were.

In the questions for overprecision regarding the 90% confidence interval the subjects had to provide a range of values indicating the minimum and maximum amount of kilocalories of the dish such that they were 90% sure that the correct answer would lie between these two numbers. This represented a width of interval estimates that depicted how the judge evaluates the own answer or own knowledge. For indicating overprecision, a range of 0 to 1 was established, whereas 0 pointed out that the right value of kilocalories of the dish was not found in the 90% confidence interval stated by the participant while 1 pointed out that the stated confidence interval did include the right value. A mean value in-between 0 and 1 referred to the corresponding level of overprecision as values towards 1 showed indication for more overprecision.

The last question concerned the third type of overconfidence - overplacement - where the subjects had to assess their own estimation performance in the corresponding round compared to other participants. In this connection, they had to choose one of eleven options ranging from ‘top 10%’ over ‘average 50%’ to ‘bottom 10%’. In the next step it was possible to compare the initial estimate given by each participant with the actual correct amount of kilocalories for the corresponding task to gain information about how close or far the participant actually was from the correct answer. Based on that, the participants were arranged in an order according to their performance in each task. These values were then compared with the participants’ initial indicated assessments of their own performance compared to others. If the result took on the value of 0, then there was no overplacement present as the subject’s guess of, for example, being among the top 20% was right. A value greater than 0 indicated some sort of overplacement and a value below 0 some sort of underplacement.

It is noteworthy that in the analyses of this paper, the trend of the judgment accuracy variable in relation to the trend of the WOA variable was chosen as a basic indicator for the optimal use of advice. Due to the restricted extent of this thesis, no focus on the actually used weighting techniques has been made.

## 5. Results

The following section includes the results of the analyses on the stated hypotheses. Further analysis results regarding demographics can be found in the appendix.

### 5.1. Hypothesis 1

To answer the first hypothesis it was analyzed whether the WOA variable would depict an increase over time and higher values in difficult tasks compared to easy tasks considering the points in time of food presentation. It is noteworthy for the latter that for 12 rounds there were six points in time as the first one indicated the first round of the easy condition and the first round of the difficult condition which also applied to the subsequent points in time.

The applied MANOVA analysis with repeated measurements showed no statistically significant effect of time and task difficulty on the WOA variable. The corresponding outcomes of the conducted analysis were F-value (5, 495) = 1.34, $p = 0.246$ ($p > .05$), partial $\eta^2 = 0.13$ which indicated a rather weak effect size. With regard to the easy condition, advice taking was highest in the first easy round with a mean value of 0.1149 and dropped immediately in the second easy round. Interestingly, the value for advice taking slightly increased in the third easy round and remained more or less constant until a new rise in the last easy round. In the difficult condition, the first round had a lower WOA value compared to the first easy round and rose in the second difficult round up to a mean value of 0.0607. As in the easy condition, the WOA values for the third, fourth and fifth difficult round remained steady although low and experienced an increase in the last difficult round. Regarding the two levels of task difficulty the observations revealed that the values of the WOA variable rather showed a decreasing tendency over the rounds. Interestingly, advice taking was higher in easy tasks than in difficult tasks which was contrary to the stated hypothesis. There is no evidence that advice taking was higher in difficult tasks compared to easy tasks as it was rather the other way round. Also it can be seen that the overall values of the advice taking variable are low, indicating that in general only few participants decided to take the advice they were shown. Due to the achieved results, the first stated hypothesis about higher values of

the WOA variable in the difficult condition was not supported. Interestingly, in the eighth round, which is the fourth round of the difficult condition, no advice at all was taken.

The results of the MANOVA analysis (mean and standard deviation in brackets) are presented in table 1. The diagram for the WOA variable is shown in figure 1 where 1 (blue line) represents the easy condition and 2 (red line) represents the difficult condition.

| | **Point of time 1** | **Point of time 2** | **Point of time 3** | **Point of time 4** | **Point of time 5** | **Point of time 6** |
|---|---|---|---|---|---|---|
| ***Easy condition*** | 0.1149 (0.5522) | 0.0135 (0.0966) | 0.0374 (0.2283) | 0.0394 (0.3098) | 0.0402 (0.2002) | 0.0540 (0.4128) |
| ***Difficult condition*** | 0.037 (0.154) | 0.0607 (0.3715) | 0.0100 (0.07882) | 0.0000 (0.0000) | 0.0035 (0.0348) | 0.0184 (0.0951) |

*Table 1: Analysis for hypothesis 1 (values for WOA variable)*

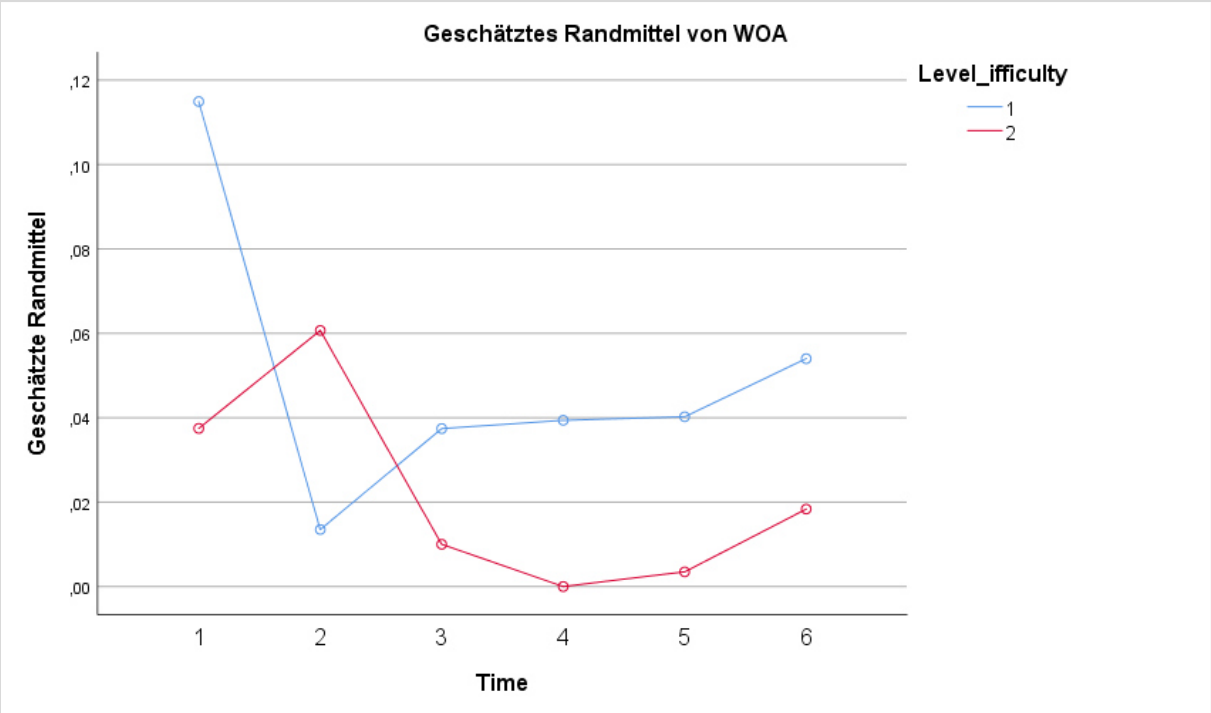



*Figure 1: Diagram for WOA*

### 5.2. Hypothesis 2

The second analysis comprised the examination whether judgment accuracy would show higher values over time in difficult tasks, considering the points in time of food presentation. The data shows a significant effect in the relationship between time and task difficulty as F-value (5, 495) = 42.42 and $p < 0.01$. The partial $\eta^2$ indicated a value of 0.3, representing a moderate effect size.

The judgement accuracy variable depicts the accuracy of the judge's final decision. Its relation to advice taking implies that advice taking and decision accuracy are positively correlated (Gardner & Berry, 1995; Sniezek et al., 2004; Yaniv, 2004b). According to the mean values shown in the table below, the absolute error of judgment accuracy rather revealed a decreasing trend in both easy and difficult tasks, referring in turn to a rising tendency of judgment accuracy. Here, until the third rounds of both conditions, the error of judgment accuracy in difficult tasks was indeed higher than in easy tasks, pointing out that participants had difficulties to state accurate estimates in difficult rounds compared to easy rounds. A slight increase of the error of judgment accuracy could be registered in the easy condition until the first four rounds before dropping in the last two rounds. Only from the fourth rounds on in both task difficulty conditions the error of judgment accuracy was higher in easy rounds than in difficult ones. This meant that participants improved their estimation process and showed a rising tendency in both conditions towards more accuracy estimates but even better estimates in difficult tasks compared to easy tasks. The value of the error of judgment accuracy even decreased rapidly in the fourth round of the difficult condition, then experienced a slight increase in the fifth difficult round before dropping again in the last difficult round. But also the easy condition exhibited a constant decline from the fourth easy round on.

As stated before, judgment accuracy is likely to be correlated with advice taking. Regarding the data of the WOA variable and the data of the error of judgment accuracy variable, it could be seen that in this study a similar behavior was partly observed. Especially in the first rounds of either task complexity condition when the WOA variable was high, the error of judgment accuracy was low and thus the

judgment accuracy was high as well (like in the first easy round). When the WOA variable was rather low, the error of judgment accuracy took on a higher value (like in the first difficult round). Interestingly, in the second difficult round, advice taking was at a moderate level but the error of judgment accuracy increased. A similar result regarding the linkage could be observed for the third difficult round with a low advice taking and a rather high error of judgment accuracy indicating a low judgment accuracy. Thereafter, the error of judgment accuracy in either complexity condition showed a declining tendency. The expectation for hypothesis 2 was that judgment accuracy would show higher values in difficult tasks than in easy tasks. Although in the beginning of the survey participants achieved more accurate results in easy tasks than in difficult tasks, in the second half of the survey judgment accuracy rose especially in the difficult condition. This implied that participants learned to state more accurate estimates over time and they did so more significantly when encountering difficult tasks. Therefore, hypothesis 2 was supported.

Table 2 shows mean values and standard deviation of the error of judgment accuracy variable considering point in time and level of task difficulty. The diagram for the error of judgment accuracy variable is shown in figure 2 where 1 (blue line) represents the easy condition and 2 (red line) represents the difficult condition.

| | **Point of time 1** | **Point of time 2** | **Point of time 3** | **Point of time 4** | **Point of time 5** | **Point of time 6** |
|---|---|---|---|---|---|---|
| ***Easy condition*** | 105.72 (109.15) | 196.48 (177.71) | 183.86 (141.73) | 256.61 (174.61) | 193.08 (124.33) | 146.11 (94.34) |
| ***Difficult condition*** | 263.81 (115.44) | 378.13 (287.14) | 315.86 (162.03) | 118.04 (81.47) | 149.52 (106.32) | 79.22 (63.11) |

*Table 2: Analysis for hypothesis 2 (values for error of judgment accuracy)*

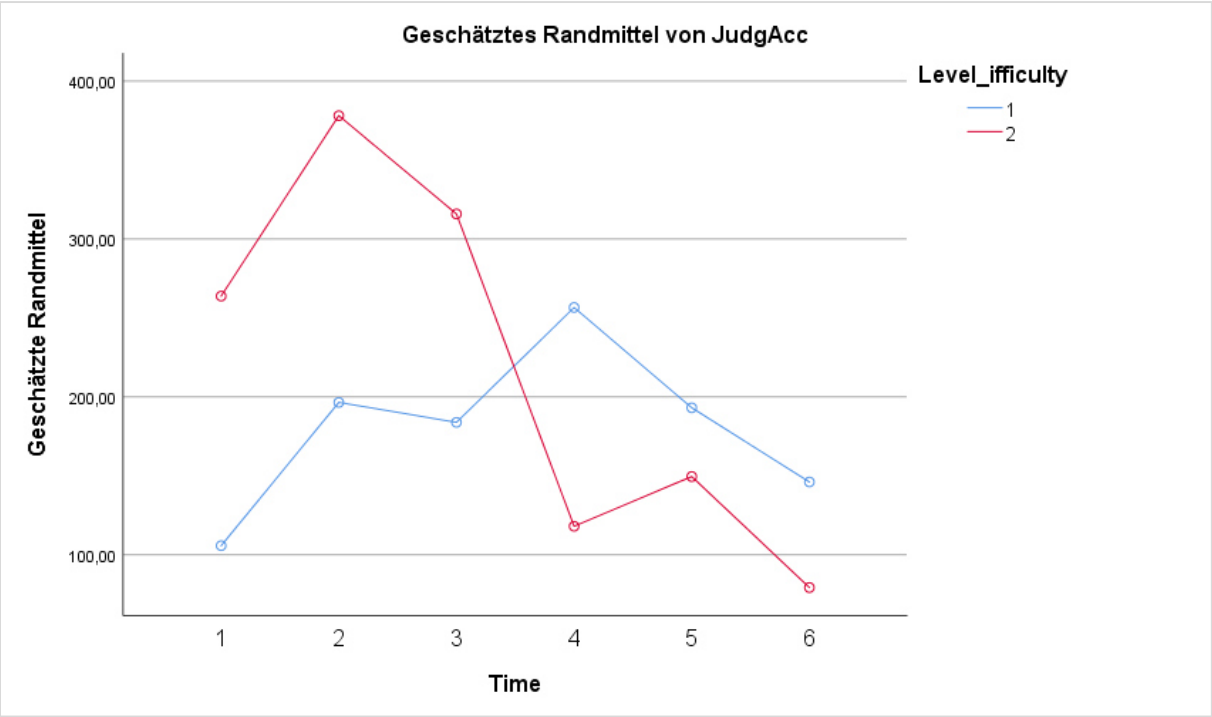



*Figure 2: Diagram for Error of Judgment Accuracy*

### 5.3. Hypothesis 3

The third analysis considered the overconfidence level of the participants where a differentiation between the three overconfidence types was carried out. These types were overestimation, overprecision and overplacement.

***Overestimation***

Having a closer look on the first type, which is overestimation, a significant effect in this case was observed as F-value (5, 495) = 27.54 and $p < .01$. In addition, the value for the partial $\eta^2$ was 0.202, representing a rather weak effect size. The data revealed that in general, the majority achieved negative values and expressed rather underestimation instead of overestimation. Yet the mean values showed a rising tendency towards overestimation in difficult tasks while the easy condition included values of a rather constant behavior. Hence, there couldn't be pronounced a total absence of overestimation as it had only a weaker presence. In the difficult condition, the phenomenon of overestimation was very weakly represented and had a severe drop in the second difficult round before rising up until a positive mean value of 13.68 in the last difficult round, indicating a presence of overestimation. Regarding the

easy condition, overestimation was slightly weak and experienced a drop in the fourth round before rising up to a value of -14.66. Hence, referring to the question about how many kilocalories participants thought their estimate was apart from the correct answer they mostly thought their answer lied more apart from the correct answer as it actually did which refers to the opposite of the tested value, namely underestimation. The values of overestimation were lower in difficult tasks than in easy tasks but only in the first half of the survey. Afterwards, participants tended to believe that their estimates were less apart from the correct answer as they actually did and thus, expressed higher overestimation. As the expectation was that overestimation will show lower values in difficult tasks compared to easy tasks over the rounds, hypothesis 3a was not supported.

Table 3 provides the mean values and the standard deviation for overestimation. The diagram for the overestimation variable is shown in figure 3 where 1 (blue line) represents the easy condition and 2 (red line) represents the difficult condition.

| | **Point of time 1** | **Point of time 2** | **Point of time 3** | **Point of time 4** | **Point of time 5** | **Point of time 6** |
|---|---|---|---|---|---|---|
| ***Easy condition*** | -17.05 (93.61) | -54.03 (197.55) | -58.03 (150.13) | -112.72 (171.27) | -46.26 (179.07) | -14.66 (142.66) |
| ***Difficult condition*** | -140.31 (174.52) | -217.97 (227.97) | -178.41 (174.20) | -34.24 (92.64) | -16.48 (140.10) | 13.68 (97.12) |

*Table 3: Analysis for hypothesis 3a (values for overestimation)*

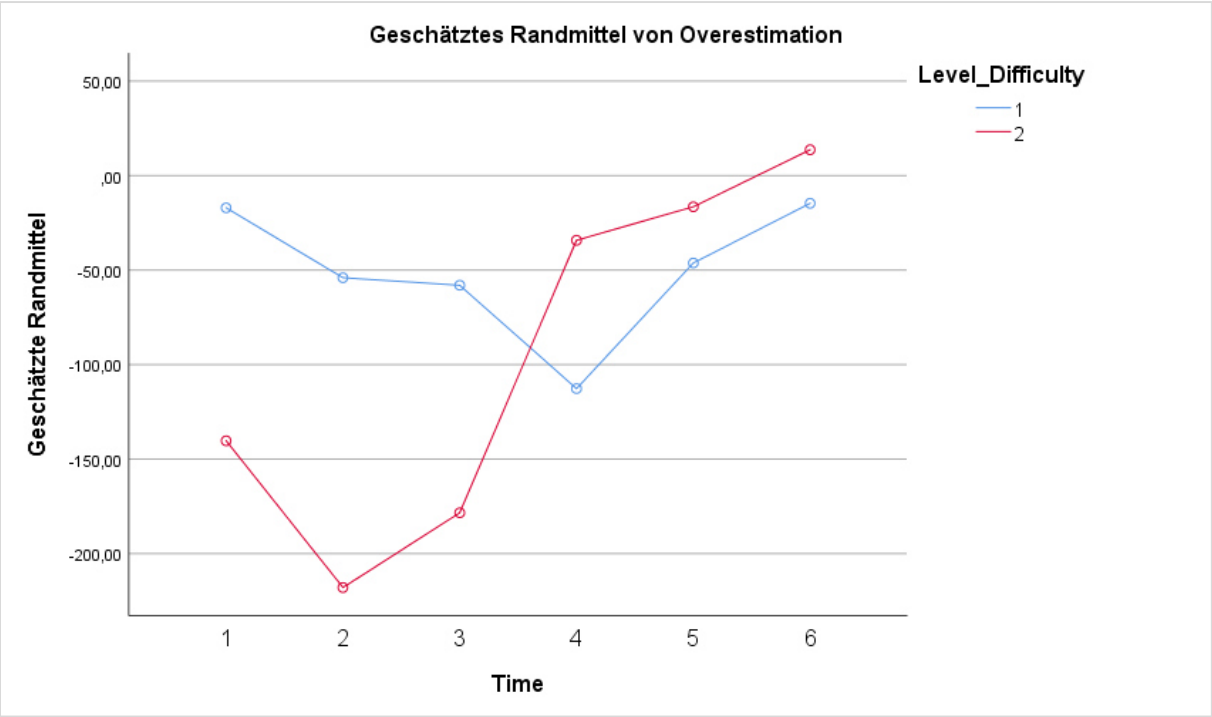



*Figure 3: Diagram for Overestimation*

***Overprecision***

An analysis on the variable overprecision also revealed a statistically significant difference as F-value (5, 495) = 10.06, $p < .01$ and partial $\eta^2 = 0.092$, pointing out a weak effect. Participants had to indicate their 90% confidence interval of which they were 90% sure that the correct answer would lie in-between. As stated before, a value of 0 indicated that the participants stated intervals that did not contain the correct answer in-between and thus, expressed signs of overprecision. In turn, a value of 1 indicated the absence of overprecision as the right answer was found in the stated boundary. Hence, every value between 0 and 1 stated a certain level of overprecision. First, a strong increase of overprecision until the fourth round of the easy condition crystallized where it experienced a harsh drop to the value of 0.27. As already stated, values towards 0 indicated high levels of overprecision and pointed out that only a low amount of intervals that were stated contained the right answer. Afterwards, the values for the variable rose straight up to 0.66 in the last easy round, indicating a rising amount of stated intervals that included the right answer in-between. Regarding the difficult condition, in the beginning, most of the confidence intervals stated by the participants in the first half did not contain the correct answer, which resulted in high levels of overprecision. This behavior changed for the last two rounds of the difficult

condition as more of the right values of the kilocalories of the dishes were found in the stated confidence intervals. It was hypothesized that participants would exhibit less overprecision in difficult tasks compared to easy tasks. The presented findings show that although overprecision was more present in difficult tasks in the first half of the survey, participants adjusted their confidence intervals and expressed less overprecision throughout the rounds. For these reasons, hypothesis 3b was supported.

The mean values and standard deviation are shown in table 4. The diagram for the overprecision variable is shown in figure 4 where 1 (blue line) represents the easy condition and 2 (red line) represents the difficult condition.

| | **Point of time 1** | **Point of time 2** | **Point of time 3** | **Point of time 4** | **Point of time 5** | **Point of time 6** |
|---|---|---|---|---|---|---|
| ***Easy condition*** | 0.61 (0.49) | 0.54 (0.50) | 0.39 (0.39) | 0.27 (0.45) | 0.47 (0.50) | 0.66 (0.48) |
| ***Difficult condition*** | 0.33 (0.47) | 0.20 (0.40) | 0.24 (0.43) | 0.23 (0.42) | 0.66 (0.48) | 0.64 (0.48) |

*Table 4: Analysis for hypothesis 3b (values for overprecision)*

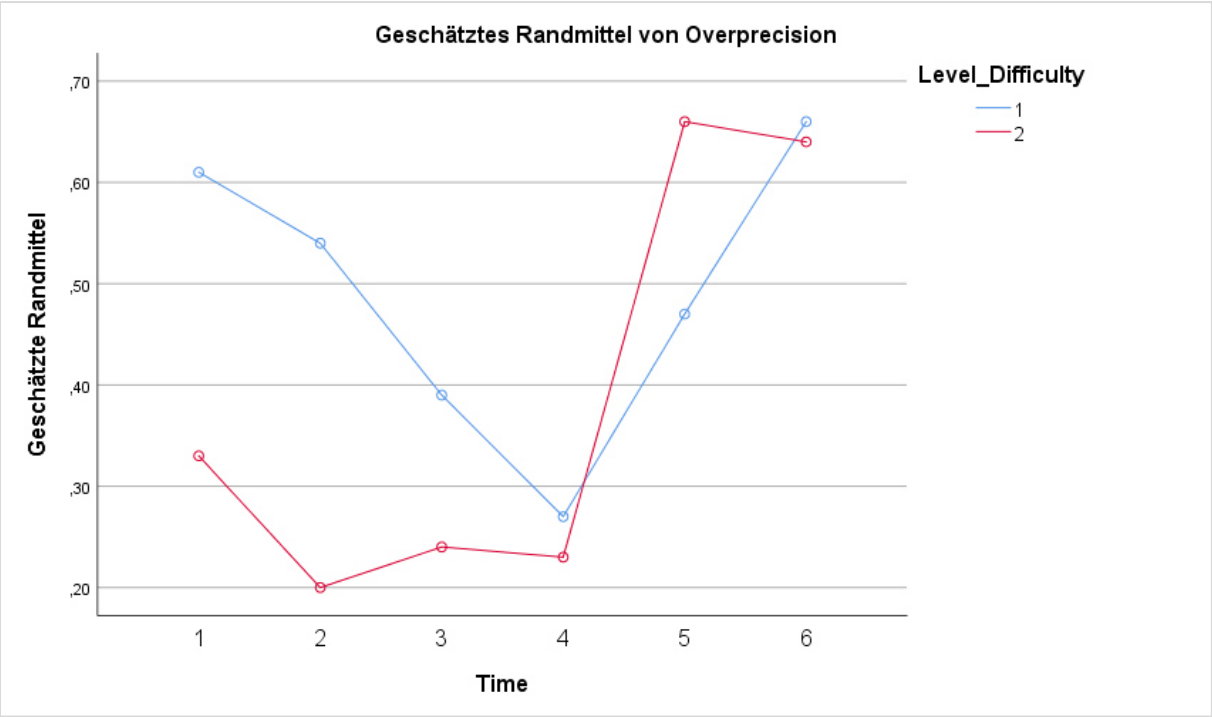



*Figure 4: Diagram Overprecision*

***Overplacement***

With regard to the third type, overplacement, no statistically significant effect was revealed as F-value (5, 495) = 0.28 and p = 0.922 (p > .05). The partial $\eta^2$ had the value of 0.003, depicting a very small effect size. To test for the participants' level of overplacement they were asked to evaluate their own performance as to how close they think their estimation is to the correct answer compared to other participants. They were presented with a scale ranging from 'top 10%' over 'average of 50%' to 'bottom 10%'. Negative values indicated a worse self-evaluation regarding the estimation performance than it actually was which refers to the phenomenon of underplacement. A positive value pointed out that the participants assessed the own estimation performance compared to other participants as better than it actually was which refers to overplacement. A value of 0 meant that the participants managed to evaluate their estimation performance adequately to their true performance. The findings revealed that participants assessed their performance higher in the beginning of the survey but lost some confidence throughout the rounds. Both task difficulty conditions showed a similar behavior. Until the third round in both conditions, the subjects tended to overplace their estimation performance compared to others. Interestingly, overplacement was lower in the first difficult round than the first easy round but higher in

the third difficult round compared to the third easy round. Despite this small rise in the difficult condition, the fourth rounds in both conditions experienced a drop in overplacement. While participants remained low values of this variable over the fifth round and showed only a slight increase in the sixth round of the difficult condition, they showed higher overplacement in the fifth round of the easy condition before dropping down to a low overplacement value of 0.37. It was expected that overplacement would be less present over the rounds especially in difficult tasks compared to easy tasks. The data showed that first, the level of overplacement was in most rounds lower in difficult tasks than in easy tasks except for the third and sixth round where it exceeded the overplacement value of the easy condition. Despite these outcomes and the drop of overplacement after the third rounds, it can't be clearly seen whether the subjects learned and managed to assess themselves in a better way as they stated answers regarding their estimation performance that were too fluctuating to announce a representative tendency. Due to these findings, hypothesis 3c was not supported.

The corresponding mean values and standard deviation are shown in table 5. The diagram for the overplacement variable is shown in figure 5 where 1 (blue line) represents the easy condition and 2 (red line) represents the difficult condition.

| | **Point of time 1** | **Point of time 2** | **Point of time 3** | **Point of time 4** | **Point of time 5** | **Point of time 6** |
|---|---|---|---|---|---|---|
| ***Easy condition*** | 1.21 (3.44) | 1.07 (3.93) | 0.95 (3.59) | 0.33 (3.85) | 0.69 (3.79) | 0.37 (3.84) |
| ***Difficult condition*** | 0.94 (3.43) | 1.02 (3.48) | 1.18 (4.21) | 0.24 (3.68) | 0.21 (4.39) | 0.41 (3.70) |

*Table 5: Analysis for hypothesis 3c (values for overplacement)*

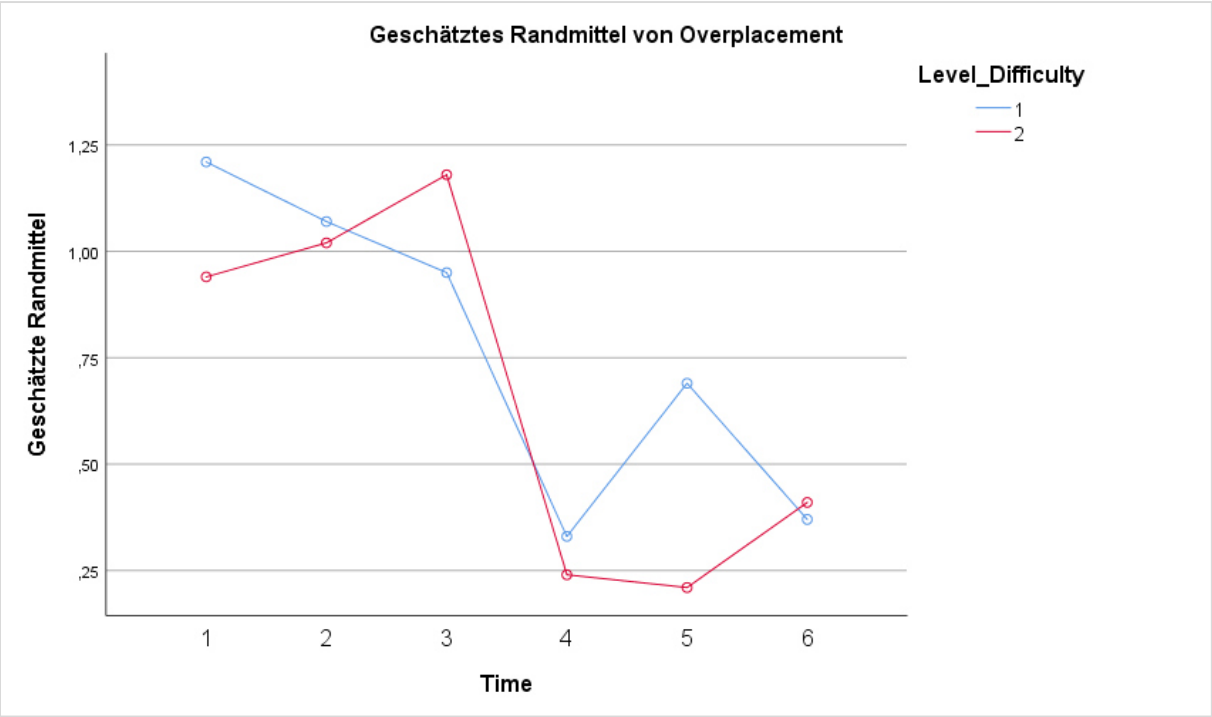



*Figure 5: Diagram Overplacement*

## 6. Discussion

### *Advice Taking*

The results show that hypothesis 1 could not be supported as the values for the WOA variable did not increase in difficult tasks as expected but rather showed a steady decreasing tendency over time. There are several possible reasons that might explain this outcome. First, participants might have taken in the first easy round the advice as it was provided by default. Since the participants did not take any practice rounds, they might have had the feeling that the advice had to be useful which would explain the high advice taking value in the first easy round. Second, after receiving feedback on the actual correct amount of kilocalories of the food and thus on their performance, participants might have decided to discount advice which would explain the decreased advice taking value in the first difficult round. In this connection, the participants might have taken into account whether the advice they had been provided with was correct or at least drifting towards the correct value or not as they saw the right answer for the task as the performance feedback. Thus, the advice provided might not have been perceived as expert advice and therefore experienced a continuous discounting. Third, the formation of a pre-advice opinion might have been likely to induce the anchoring effect (Tversky & Kahneman, 1974). Fourth, some although very few people stated in the comments section that they did not know some certain dishes at all. This can be applied on the last two rounds which corresponded to the achieved data as in the respective last round of each difficulty condition the WOA values slightly increased. Fifth, regarding the last two questions in the survey that addressed the participants general attitude towards kilocalories in daily life, 56.8% stated that they don't think about kilocalories of their food and 71.8% ascribed no role for kilocalories in their daily routine. Moreover, participants might have been confused by the pictures of the dishes as they were shown a full plate of the dish while the tasks and ingredients lists referred to kilocalories per serving. Thus, it can be assumed that most participants were not used to the thinking process of paying attention to or calculating kilocalories of food and thus, had no proper idea of how to assess these factors. The reason for no advice taking in the fourth round of the difficult condition can likely be explained by the type of the dish chosen for the task. As it showed a simple dish with only a few ingredients, participants might have had the feeling of being able to solve the task on

their own although they did not have information about the exact ingredients which would explain the corresponding tendency towards overestimation and the high value of overprecision for that round.

### *Judgment Accuracy*

In the matter of judgment accuracy, the stated hypothesis was supported as the data showed an increase in the accuracy of the estimates stated by the participants over the rounds. As participants improved their decision accuracy especially in the difficult condition of the second half of the survey, there are still some influential factors that might have impacted this outcome. On the one hand, unknown dishes or ingredients of the food might have led to poor performance regarding the judgment accuracy in the first few difficult rounds. And also, another factor is that the variety of dishes might have led to problems in the participants' judgments. Although the easy tasks included an exact ingredients list and all dishes represented food that was commonly known some people might have had difficulties to estimate the kilocalories as they were not used to it. On the other hand, the increase in judgment accuracy cannot only be justified by the participants' learning progress due to the performance feedback. There is the possibility that, as the study was conducted online, some participants might have come up with the idea to search for some ingredients or recipes of the food on the internet especially when being confronted with dishes in the difficult condition. Also, it is not unlikely that participants might have been influenced by their environmental surroundings as they might have had the opportunity to consult other people like family members or friends in order to state their estimates. If participants decided to ask other people in their environmental surroundings, this, too, would indicate some sort of advice taking but it was not possible to examine this phenomenon in the online study. Still, it is likely that the subjects learned to give more accurate estimates due to the performance feedback they received. Despite findings showed that judgment accuracy did not always align with the participants' advice taking behavior, nonetheless there are signs of a learning progress as the participants' estimates became more accurate during the last rounds and especially in the difficult condition.

### *Overestimation*

Although there is no proof for the total absence of overestimation, the testing results showed rather signs of underestimation and less for overestimation in the first half of the survey for both task difficulty conditions wherefore the stated hypothesis was not supported. Reasons for this outcome in the difficult condition might be linked to the elevated value of WOA. In general, when advice taking was high, the level of overestimation decreased, so as especially in the second difficult round. Interestingly, the values of the next rounds showed a tendency towards overestimation but still remained in the field of underestimation. A very plausible reason for this outcome is the participants' unfamiliarity with the indication process of the own estimation performance. Still, in the second half of the survey, there was less underestimation but rather an increase up to a positive value of overestimation in the sixth difficult round. A possible justification for this trend can be provided by the linkage to judgment accuracy. According to the data, as the error of judgment accuracy decreased and thus, the judgment accuracy increased, so did the values for overestimation. Participants might have felt more secure as the performance feedback provided might have informed them about their estimates being accurate. Subsequently in the next rounds, the subjects believed their initial estimates to be less kilocalories apart from the correct answer than they actually were, resulting in higher values of overestimation. Hence, their behavior drifted away from underestimation. Also, the participation in the survey was anonymously and there was no punishment which might have induced a higher overestimation as the participants did not risk to lose their stake.

### *Overprecision*

Referring to the second type of overconfidence, the stated hypothesis was supported. In the first half of the survey the variable of overprecision took on almost constant values in the difficult condition while it showed an increasing tendency in the easy rounds. This can be justified with a conglomerate of the aforementioned reasons. As most participants indicated confidence intervals in the first easy round that contained the correct value of the dish they might have felt more confident in the next rounds and continued to indicate confidence intervals more narrow, at the same time increasing the chance that the

correct value was not in-between these boundaries. This trend towards overprecision reached its peak in the fourth round for each of the difficulty conditions with values of 0.27 for the easy task and 0.23 for the difficult task. After that, the level of overprecision dropped for both conditions which might be due to the experience of the participants. By receiving performance feedback after each round they had information on how well their estimates were chosen. But as the data from the analysis might indicate that the intervals in the following rounds were more precise as a majority contained the correct answer, it is more likely that the participants decided to state wider intervals to make sure that the correct amount of kilocalories of the dish was definitely found in the stated boundaries. In addition, having a closer look on the tasks that were presented in the second difficult round and the fourth rounds of both the easy and difficult condition it is likely that participants felt too secure in their estimates as the dishes shown in the pictures consisted of only few ingredients and might have appeared more common to them. These results can also be linked to the findings of hypothesis 1 as the data showed low advice taking especially in the fourth difficult round.

***Overplacement***

In terms of overplacement, the last of the discussed types of overconfidence, the data showed high values for both conditions until the third rounds which implied that the subjects considered their estimation performance as better than it actually was. Although the judgment accuracy variable in the difficult condition showed low values for the first half of the survey, overplacement in these rounds was still high. For the next rounds the data depicted a drop which pointed out the learning process of the participants. By receiving feedback on their performance, participants had information also on their judgment accuracy. It is assumed that they therefore learned to assess their estimation performance compared to other participants in a more realistic way. This implied lower levels of overplacement which are shown in the data of the analysis for this hypothesis. Still, due to fluctuations in the trend development, the stated hypothesis was not supported.

## 7. Conclusion

This work started with a literature review on the included variables which are advice taking, judgment accuracy, overconfidence, task difficulty and performance feedback. Most prior research in the field of judgment and decision making put focus on the one hand on individual choices without the impact of others' opinions and on the other hand on the decisions reached by groups in a consent. In fact, most decisions made are a mixture of these two phenomena. People are seldom aware of it but most of the time decisions are made under external influence such as advice, additional information or expertise, or social support (Van Swol et al., 2018). The role of advice in the decision-making process is essential as it supports the judge in receiving, considering and using additional information and knowledge (Van Swol et al., 2018). The here presented procedure of estimation tasks under variation of task difficulty represents a helpful tool in experimental settings for the examination of the relations between the aforementioned variables.

This paper comprises a proper overview of the literature on advice, judgment accuracy, overconfidence, task difficulty and performance feedback. As to the advice section, the different types of advice, the research on advice utilization and advice discounting as well as other influencing factors like the perceived power and expertise of the agents or the judge's willingness to pay for advice were introduced. Furthermore, existing findings regarding the task difficulty were confirmed as it was found that the difficulty condition affects the judge's advice taking behavior to the extent as the decision maker is more willing to accept the received advice when facing difficult tasks compared to easy tasks (Gino & Moore, 2007; Schrah et al., 2006). Although prior research has proven the existence of a connection between confidence and judgment accuracy, this relationship succumbs to a weak persistence and the advice giver often tends to exhibit some level of overconfidence (Erev et al., 1994; Van Swol & Sniezek, 2005). Other researchers found that the constant receipt of performance feedback is likely to enhance the learning goal orientation of participants as subjects underlie the norm of aiming for learning progress (Ghoshal & Bartlett, 2007). Individuals aim for social and cognitive interaction as they try to find a new way to illustrate their decision problem (Yates et al., 1996; Zarnoth & Sniezek, 1997) and to realize an

improved final decision (Gino et al., 2009; Harvey & Fischer, 1997; Yaniv et al., 2011). The so-called Judge-Advisor System (JAS, see Sniezek & Buckley, 1995; Sniezek & Van Swol, 2001) represents a very common phenomenon in the research area of advice taking. Not only the characteristics of decision maker and advice giver are influencing advice taking behavior but also the weight and integration of the decision maker's initial opinion with the recommendation given by the advisor. Prior research has found that usually individuals tend to underweight the recommendation they had been given (Yaniv, 2004b; Yaniv & Kleinberger, 2000). These findings justify the essential purpose of advice taking in information processing (Brooks et al., 2015).

After the presentation of the hypotheses that were to be examined accompanied with the corresponding justifications the methodology for the conducted research was presented including information about the procedure, the participants and the data measurement description. 103 out of 183 participants finished the online survey where they had to estimate the amounts of kilocalories of dishes shown in pictures and to answer some questions regarding their level of the overconfidence types which are overestimation, overprecision and overplacement. Furthermore, participants were provided with an advice which represented the estimates from a randomly chosen individual, followed by the opportunity to adjust the initial estimate and the provision of the performance feedback in form of information on the actual amount of kilocalories of the corresponding dish. Subsequently, analyses on the impact of task difficulty and performance feedback on the use of advice, judgment accuracy and the types of overconfidence were conducted to answer the research question whether participants learn to use advice in a more optimal way over the rounds when they receive performance feedback. Although the first easy round revealed the highest advice taking value, the WOA variable was in general very low and this phenomenon was observed over all rounds. Thus, hypothesis 1 was not supported. With regard to the analysis on judgment accuracy, the data showed that participants were indeed able to learn to state more accurate estimates over time and to do so more in difficult tasks than in easy tasks, such that the second hypothesis was supported. Next, analyses on the three types of overconfidence were conducted. With regard to the first one, which is overestimation, the findings showed a statistically significant effect but the participants rather expressed behavior of underestimation and only showed slight tendencies of

overestimation in the last rounds. Thus, this hypothesis was not supported as the variable was expected to show lower values in difficult tasks. A significant effect was also detected regarding the variable of overprecision. In the easy condition, high values of overprecision were present. As first, more overprecision was expressed in the first difficult rounds, the values decreased afterwards which might be justified with the type of dish shown in the pictures such that there was less overprecision in difficult tasks than in easy tasks and the hypothesis was supported. With reference to the last type of overconfidence, the hypothesis was not supported as the values that resulted from the data analysis were too fluctuating over the rounds to show a proper trend development.

The aim of this paper was to examine whether participants learn to use the advice they have been provided with in a more optimal way. Conclusively, it can be said that the answer to this question is difficult to provide as the overall advice taking behavior over the rounds was lower than expected and in the eighth round which was the fourth difficult round even no advice at all was taken by the participants. The analysis of the advice taking variable did not achieve the results that were mentioned in several studies by other researchers, namely that advice taking is higher in difficult tasks than in easy tasks, which might have depended on the variety of dishes shown in the pictures or the knowledge of the participants regarding kilocalories of food. As a slight increase of the WOA variable only appeared in the latter rounds, it is unlikely that this was due to the provision of performance feedback as the participants should have reacted to it in earlier stages of the survey. The findings revealed that significant results were achieved in terms of the impact of task difficulty. Although advice taking did slightly increase in the last rounds of both task difficulty conditions and in this connection, judgment accuracy took on high values in these rounds, it can't be stated for sure that judgement accuracy is positively correlated with advice taking. This would require a regression analysis which was not included in this study due to a different focus. Still, results revealed that judgment accuracy which was selected to be an indicator for the use of advice increased over the rounds. This phenomenon was observed even stronger in the difficult tasks compared to easy tasks. Regarding judgment accuracy, more accurate estimates were given in difficult tasks compared to the easy ones. Also overestimation was barely present as the participants mostly expressed underestimation. This can be justified with the unfamiliarity with the task

of indicating the amount of kilocalories being apart from the correct answer. Although the data revealed that participants were more careful in their estimates about the kilocalories being apart when they encountered difficult tasks and felt more safe to state their guesses in easy tasks, these tendencies changed such that the first positive value indicating overestimation appeared in the last round of the difficult condition. In the beginning, levels of overprecision were higher in the complex condition and decreased over the rounds stronger for difficult tasks. This might be due to two reasons. First, the received feedback performance might have led participants to learn to assess their confidence intervals more precise or second, that they arranged their confidence intervals wider to increase the chance that the correct value would fall in-between the stated boundaries. Nonetheless, both alternatives were of some benefit in cases where participants received information about their feedback performance. Thus, this variable indeed had an impact on overprecision as it helped participants to assess their confidence intervals. Although levels of overplacement decreased in the course of the survey for both task difficulty conditions despite one quick jump in the fifth round of the easy condition, this variable showed too many fluctuations such that no clear argumentation about its susceptibility to task difficulty can be stated. Still, performance feedback might have influenced the participants as they indicated less signs of overplacement in the second half of the survey. In summary it can be said that task difficulty and performance feedback did have an effect on the mentioned variables as in some cases participants learned to assess their performance in a better way and also to state more accurate estimates. But too many influencing factors were present and the experimental period of this study might have been not long enough to assess the long-term behavior of the participants. Thus, a more expanded study design is necessary to examine the effect in a more proper way and to identify the participants' learning behavior in the advice-taking process.

In conclusion it can be said that the examination procedure of only the judge's weighting behavior towards the own decision and the opinion received from the advisor has to be expanded as, for instance, this present study only included the advisor's recommendations given by default and suspended the idea of the judge being able to choose whether or not to solicit the advice. In this case, a comprehensive theory on the matter of advice utilization is beneficial as it might provide explanations regarding the

environmental conditions, the times and the individuals that are most suitable to be sought for advice by the decision makers. Furthermore, this comprehensive theory should be able to give information about the weighting techniques of decision makers towards the received recommendations as well as the judges' behavior in settings with unsolicited advice.

### 7.1. Limitations

Although the measures and outcomes of the study presented in this paper are of adequate quality, they do contain some limitations which will be discussed in this chapter.

An obvious limitation that stands out when having a closer look on the data of the questionnaire right after closing the survey is the participants' comprehension of the questions. Here, some people did not understand how to properly state the upper and lower boundaries for the 90% confidence interval as their chosen intervals did not even contain their own initial estimate. Therefore, these cases had to be corrected as the data was not usable for further data processing. Moreover, many participants did start the survey but changed their mind and cancelled their participation. Hence, this incomplete data could not be used either. Also, it could be observed that some participants did not pay attention to honest answers but wanted to finish the survey anyways and entered the same value for every upcoming question. All these mentioned cases induced a high dropout rate of 43.72% and delayed the completion of the survey phase due to an ongoing search for participants. In addition, as the survey was presented to the participants online, it can't be ruled out that people answered the survey with the help of others or were influenced by various environmental settings or their emotional state. Besides, the provision of participants with the opportunity to do some test rounds could have increased the subjects' performance but also their level of overconfidence. This can be led back to the theory of the *illusions of learning* which was introduced by several researchers (Adams & Goetz, 1973; Harvey, 1994; Marteau et al., 1989). This phenomenon appears when participants believe that the experience they achieve in trial rounds is an indicator for their judging expertise, or more precisely, when participants overweight the

effectiveness of trial rounds. On the other hand, with the help of trial rounds participants can get used to the thinking process which might have supported their decision process in this study.

In addition, the method used in this paper to examine the advice taking has still some flaws. In cases where the judge's initial opinion takes the same value as the advice, the WOA variable has undefined results as the denominator equals zero. Also, in cases where the judge's final estimate takes the same value as the judge's initial opinion, the WOA variable is zero which means that there is no advice taking. However, with the denominator drifting towards a very small value it is possible that the decision maker might have accepted the recommendation as the own initial estimate appeared to be already close enough to the correct value and had more confidence in the provided recommendation. In contrast, receiving advice does not always result in the adjustment of one's own estimate but also in the adjustment of the judge's confidence. Moreover, a smaller distance between the received advice and the judge's initial estimate might decrease the confidence interval. Another point regarding the limitations of the WOA variable is that this method does not make any difference between situations in which movements (i.e. *hold* and *shift*) of the judge's final opinion towards or away from the advice occur whereas the occurrence of the latter is rare. In addition, besides having a lower bound of zero, this measurement has no upper bound which means that, although this occurrence is seldom, the WOA variable takes the value greater than one when the judge's final opinion exceeds the value of the advice (Gino, 2008; Harvey & Fischer, 1997). In this connection, there was no consideration of whether the judge's initial estimate, the advice and the judge's final estimate represented the same value or whether the judge chooses a different one as there was no focus on further investigation on the *hold* or *shift* behavior presented by Bonaccio and Dalal (2006). As well, no closer look on the presented phenomena of the judge's final estimate *approaching*, *overshooting* or *moving away* from the advice was included (Bonaccio & Dalal, 2006).

Moreover, according to the literature, the general assumption states that judge and advisor possess the same knowledge and expertise. Hence, a WOA of 0.5 would depict the optimal value for judgment accuracy but this concept can't be applied onto the present paper. Unlike the study by Gino and Moore (2007) who incorporated advice that gave hints on the correct answer the advice in this paper is retrieved

from a randomly chosen participant who worked on the estimation tasks beforehand. This advice did not necessarily represent the correct values of the kilocalories of the dishes and thus, cannot be interpreted as hints towards the correct answer of the task as they only represent the individual opinion of another person. Referring to this, a problem can evolve when the participants rely on the advice and exhibit a WOA value of 1, meaning that the advice was fully accepted. But this does not induce a higher judgment accuracy since the advice is not mandatory correct.

Also, other influencing variables could have affected the outcome of this study when included. For example, considering the traits of the advisor such as his or her level of confidence. Advice discounting is less common when the advisor has a higher level of expertise (Goldsmith & Fitch, 1997; Harvey & Fischer, 1997; Sniezek et al., 2004) or is older, seems wiser, more educated or experienced (Feng & MacGeorge, 2006). In addition, the confidence of the advisor plays a role in the advice taking process as recommendations from advisors expressing themselves as confident are less discounted and valued more compared to advisors with lower confidence (Sniezek & Buckley, 1995; Sniezek & Van Swol, 2001; Van Swol & Sniezek, 2005). But in the present survey, the decision makers did not have any information about the advisor's expertise or confidence.

Moreover, incentives are known to affect the participants' motivation like rewards for the best outcome or punishments for the worst (Sniezek et al., 2004). In the same way, this paper took account of the case that participants had the option to form a pre-advice opinion, namely the initial estimate, and had the chance to revise it (Rader et al., 2015) but they did not have the opportunity to state whether or not to receive the advice as it was obligatory (Gino & Moore, 2007).

Furthermore, the study conducted in this paper did not represent real-world settings as it was a questionnaire provided online. First, there were no personal interactions between judge and advisor as it would be more common in the real world. Second, the task implemented in the survey did not represent a task most people are confronted with in their everyday life. Thus, they were not used to the corresponding thinking process which might have yielded unclear estimates.

### 7.2. Future Research Approach

As research on traits of both judge and advisor are limited so far, a closer look on that topic would contribute to the psychological processes in the JAS. Furthermore, more suitable for the depiction of real-world decisions are for example tasks regarding purchase decisions of everyday products or financial investments under consideration of risk averse and risk tolerant participants.

In addition, advice utilization is influenced by multiple social factors that could not have been considered in this paper so as to the individual's feeling of pressure or urge to use the received advice. This might be the case also if the advice giver appears to have less expertise or knowledge (Gino et al., 2012; Harvey & Fischer, 1997; Yaniv & Kleinberger, 2000) and even if the decision maker knows about the uselessness or faultiness of the advice (Schultze et al., 2017). Decision makers might succumb to social pressure in the subject matter of advice taking and utilization (Harvey & Fischer, 1997) such as in situations where the advice is considered a gift and its rejection would appear to be impolite (Cialdini, 1984). As there is barely research on the social influencing factors in the context of decision-making processes (Payne et al., 1988), it is useful to do further research on that topic by including the emotional state of the involved agents in the advice-taking process and to control the environmental settings of the experiment, considering all influencing factors. In addition, although Maister et al. (2001) have found trust to be the most important factor in advice utilization, not enough additional research on that subject matter has centered the relationship between the judge and the advisor. It is proposed to start further investigation over a longer period of time. This can shed some light on that issue as trust relationships take time to develop and can't be studied in sporadic experimental settings.

In the matter of pre-advice opinions, another interesting research focus can be set on the examination of the adjustment behavior of advisors in their advising process. Therefore, experimental settings can offer to the advisor the possibility to know about the judge's opinion made before giving the advice and then to examine whether the advisor maintains the recommendation or chooses to adjust it before offering it

to the judge. Also, to focus on the development of the advisor's pre-advice confidence compared to the post-advice confidence can provide further insight into the topic.

Moreover, as the present study did not provide any further details on the weighting strategies applied by the participants, this expresses an interesting point for further research in the field of the optimal use of advice. As well, investigation on the behavior of the judge's initial estimate would bring more insight into the area of advice taking, considering movements according to the hold and shift behavior or whether the final estimate approaches, overshoots or moves away from the advice (Bonaccio & Dalal, 2006).

Furthermore, the present study presented only some forms of advice that are until now perceived as common in the literature of advice. With regard to the expounded range of advice definitions it is likely that novel ways of measurement of advice utilization and discounting have to be established as well. It can be beneficial to do research on other alternative types of advice in order to investigate differences in the choice regarding the receipt and provision of those different forms. Goldsmith (2000) argued that the temporal order of the advice can play an important role as well regarding the extent of the judge's advice utilization. For instance, a judge might be more susceptible to a specific advice from the advisor after receiving emotional support. As emotional support seems to be the most commonly given type of advice (Gibbons et al., 2003), advice discounting may occur in this case much less compared to other types of advice. Thus, a proper investigation on the variety of types of advice is needed.

An advantageous approach would be to see this paper as a basis for further research, to expand the research design for a longer period under consideration and to adjust the subject pool for example by involving participants that have knowledge in the field of kilocalories. The data of the analyses showed that most of the tendencies evolved after the first half of the survey. Therefore, it is useful to establish an experimental design to examine these variables and their relationships for longer periods to gain more information about their trend development. Also, a comparison of this research with other studies that did not contain the variable of performance feedback might provide a closer insight into its influence.

## Bibliography


Adams, J. A., & Goetz, E. T. (1973). Feedback and Practice as Variables in Error Detection and Correction. *Journal of Motor Behavior*, *5*(4), 217–224. https://doi.org/10.1080/00222895.1973.10734968

Advice. (2020). Cambridge International Dictionary of English. https://dictionary.cambridge.org/dictionary/english/advice

Alicke, M. D., & Govorun, O. (2005). *The Better-Than-Average Effect*.

Alicke, M. D., Vredenburg, D. S., Hiatt, M., & Govorun, O. (2001). The "Better Than Myself Effect." *Motivation and Emotion*, *25*(1), 7–22. https://doi.org/10.1023/A:1010655705069

Annosi, M. C., Monti, A., & Martini, A. (2020). Individual learning goal orientations in self-managed team-based organizations: A study on individual and contextual variables. *Creativity and Innovation Management*, caim.12377. https://doi.org/10.1111/caim.12377

Armstrong, J. S. (2001). *Principles of forecasting: A handbook for researchers and practitioners*. Netherlands: Kluwer Academic Publishers.

Ashton, A. H., & Ashton, R. H. (1985). Aggregating Subjective Forecasts: Some Empirical Results. *Management Science*, *31*(12), 1499–1508. https://doi.org/10.1287/mnsc.31.12.1499

Bar-Hillel, M. (2001). Subjective probability judgments. In *International Encyclopedia of the Social & Behavioral Sciences* (pp. 15247–15251). Elsevier Science Ltd.

Baron, R. A. (2000). *Psychological Perspectives on Entrepreneurship Cognitive and Social Factors in Entrepreneurs' Success*. ResearchGate. https://www.researchgate.net/publication/258128226_Psychological_Perspectives_on_Entrepreneurship_Cognitive_and_Social_Factors_in_Entrepreneurs'_Success

Bolger, F., & Önkal-Atay, D. (2004). The effects of feedback on judgmental interval predictions. *International Journal of Forecasting*, *20*(1), 29–39. https://doi.org/10.1016/S0169-2070(03)00009-8

Bonaccio, S., & Dalal, R. S. (2006). Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences. *Organizational Behavior and Human Decision Processes*, *101*(2), 127–151. https://doi.org/10.1016/j.obhdp.2006.07.001

Brehmer, B., & Hagafors, R. (1986). Use of experts in complex decision making: A paradigm for the study of staff work. *Organizational Behavior and Human Decision Processes*, *38*(2), 181–195. https://doi.org/10.1016/0749-5978(86)90015-4

Brooks, A. W., Gino, F., & Schweitzer, M. E. (2015). Smart People Ask for (My) Advice: Seeking Advice Boosts Perceptions of Competence. *Management Science*, *61*(6), 1421–1435. https://doi.org/10.1287/mnsc.2014.2054

Brown, J. D. (2012). Understanding the Better Than Average Effect: Motives (Still) Matter. *Personality and Social Psychology Bulletin*, *38*(2), 209–219. https://doi.org/10.1177/0146167211432763

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, *62*(3), 193–217. https://doi.org/10.1037/h0047470

Brunswik, E. (1956). Psychology and the Representative Design of Experiments. *University of California Press*, *2nd ed.*

Budescu, D. V., & Rantilla, A. K. (2000). Confidence in aggregation of expert opinions. *Acta Psychologica*, *104*(3), 371–398. https://doi.org/10.1016/S0001-6918(00)00037-8

Budescu, D. V., Rantilla, A. K., Yu, H.-T., & Karelitz, T. M. (2003). The effects of asymmetry among advisors on the aggregation of their opinions. *Organizational Behavior and Human Decision Processes*, *90*(1), 178–194. https://doi.org/10.1016/S0749-5978(02)00516-2

Campbell, W. K., Goodie, A. S., & Foster, J. D. (2004). Narcissism, confidence, and risk attitude. *Journal of Behavioral Decision Making*, *17*(4), 297–311. https://doi.org/10.1002/bdm.475

Cesarini, D., Sandewall, Ö., & Johannesson, M. (2006). *Confidence interval estimation tasks and the economics of overconfidence*.

Cialdini, R. B. (1984). *Influence: The Psychology of Persuasion*. NY: William Morow & Co.

Conger, R. F., & Wolstein, Ch. R. (2004). *Managing overconfidence in pricing*. *Emphasis*(Vol. 2), 10–13.

Cooper, R. S. (1991). Information processing in the judge-adviser system of group decision-making. *Unpublished Master's Thesis, University of Illinois, Urbana-Champaign*.

Cross, R., Borgatti, S. P., & Parker, A. (2001). Beyond answers: Dimensions of the advice network. *Social Networks*, *23*(3), 215–235. https://doi.org/10.1016/S0378-8733(01)00041-7

Dalal, R. S., & Bonaccio, S. (2010). What types of advice do decision-makers prefer? *Organizational Behavior and Human Decision Processes*, *112*(1), 11–23. https://doi.org/10.1016/j.obhdp.2009.11.007

de Hooge, I. E., Verlegh, P. W. J., & Tzioti, S. C. (2014). Emotions in Advice Taking: The Roles of Agency and Valence: Emotions in Advice Taking. *Journal of Behavioral Decision Making*, *27*(3), 246–258. https://doi.org/10.1002/bdm.1801

Erev, I., Wallsten, T. S., & Budescu, D. V. (1994). Simultaneous over- and underconfidence: The role of error in judgment processes. *Psychological Review*, *101*(3), 519–527. https://doi.org/10.1037/0033-295X.101.3.519

Fast, N. J., Sivanathan, N., Mayer, N. D., & Galinsky, A. D. (2012). Power and overconfident decision-making. *Organizational Behavior and Human Decision Processes*, *117*(2), 249–260. https://doi.org/10.1016/j.obhdp.2011.11.009

Feng, B., & MacGeorge, E. L. (2006). Predicting Receptiveness to Advice: Characteristics of the Problem, the Advice-Giver, and the Recipient. *Southern Communication Journal*, *71*(1), 67–85. https://doi.org/10.1080/10417940500503548

Fischer, I., & Harvey, N. (1999). Combining forecasts: What information do judges need to outperform the simple average? *International Journal of Forecasting*, *15*(3), 227–246. https://doi.org/10.1016/S0169-2070(98)00073-9

Fleming, S. M., & Daw, N. D. (2017). Self-evaluation of decision-making: A general Bayesian framework for metacognitive computation. *Psychological Review*, *124*(1), 91–114. https://doi.org/10.1037/rev0000045

Galton, F. (1907). Vox Populi. *Nature*, *75*(1949), 450–451. https://doi.org/10.1038/075450a0

Gardner, P. H., & Berry, D. C. (1995). The effect of different forms of advice on the control of a simulated complex system. *Applied Cognitive Psychology*, *9*(7), S55–S79. https://doi.org/10.1002/acp.2350090706

Gerrish, E. (2016). The Impact of Performance Management on Performance in Public Organizations: A Meta-Analysis. *Public Administration Review*, *76*(1), 48–66. https://doi.org/10.1111/puar.12433

Ghoshal, S., & Bartlett, C. A. (2007). Linking organizational context and managerial action: The dimensions of quality of management. *Strategic Management Journal*, *15*(S2), 91–112. https://doi.org/10.1002/smj.4250151007

Gibbons, A. M., Sniezek, J. A., & Dalal, R. S. (2003, November). *Antecedents and consequences of unsolicited versus explicitly solicited advice*. Symposium presented at the annual meeting of the Society for Judgment and Decision Making, Vancouver, BC.

Gillenkirch, P. D. R. (2018, February 19). *Definition: Overconfidence* [Text]. https://wirtschaftslexikon.gabler.de/definition/overconfidence-53937; Springer Fachmedien Wiesbaden GmbH. https://wirtschaftslexikon.gabler.de/definition/overconfidence-53937/version-276999

Gino, F. (2008). Do we listen to advice just because we paid for it? The impact of advice cost on its use. *Organizational Behavior and*

*Human Decision Processes*, *107*(2), 234–245. https://doi.org/10.1016/j.obhdp.2008.03.001

Gino, F., Brooks, A. W., & Schweitzer, M. E. (2012). Anxiety, advice, and the ability to discern: Feeling anxious motivates individuals to seek and use advice. *Journal of Personality and Social Psychology*, *102*(3), 497–512. https://doi.org/10.1037/a0026413

Gino, F., & Moore, D. A. (2007). Effects of task difficulty on use of advice. *Journal of Behavioral Decision Making*, *20*(1), 21–35. https://doi.org/10.1002/bdm.539

Gino, F., & Schweitzer, M. E. (2008). Blinded by anger or feeling the love: How emotions influence advice taking. *Journal of Applied Psychology*, *93*(5), 1165–1173. https://doi.org/10.1037/0021-9010.93.5.1165

Gino, F., Shang, J., & Croson, R. (2009). The impact of information from similar or different advisors on judgment. *Organizational Behavior and Human Decision Processes*, *108*(2), 287–302. https://doi.org/10.1016/j.obhdp.2008.08.002

Gino, F., Sharek, Z., & Moore, D. A. (2011). Keeping the illusion of control under control: Ceilings, floors, and imperfect calibration. *Organizational Behavior and Human Decision Processes*, *114*(2), 104–114. https://doi.org/10.1016/j.obhdp.2010.10.002

Godek, J., & Murray, K. B. (2008). Willingness to pay for advice: The role of rational and experiential processing. *Organizational Behavior and Human Decision Processes*, *106*(1), 77–87. https://doi.org/10.1016/j.obhdp.2007.10.002

Goldsmith, D. J. (2000). Soliciting advice: The role of sequential placement in mitigating face threat. *Communication Monographs*, *67*(1), 1–19. https://doi.org/10.1080/03637750009376492

Goldsmith, D. J., & Fitch, K. (1997). The Normative Context of Advice as Social Support. *Human Communication Research*, *23*(4), 454–476. https://doi.org/10.1111/j.1468-2958.1997.tb00406.x

Harries, C., Yaniv, I., & Harvey, N. (2004). Combining advice: The weight of a dissenting opinion in the consensus. *Journal of Behavioral Decision Making*, *17*(5), 333–348. https://doi.org/10.1002/bdm.474

Harvey, N. (1994). Relations between confidence and skilled performance. In *Subjective probability* (pp. 321–352). John Wiley & Sons.

Harvey, N., & Fischer, I. (1997). Taking Advice: Accepting Help, Improving Judgment, and Sharing Responsibility. *Organizational Behavior and Human Decision Processes*, *70*(2), 117–133. https://doi.org/10.1006/obhd.1997.2697

Harvey, N., & Fischer, I. (2005). Development of experience-based judgement and decision making: The role of outcome feedback. *The Routines of Decision Making*, *In T. betsch&S. Haberstroh (Eds.)*, 119–137.

Harvey, N., & Harries, C. (2004). Effects of judges' forecasting on their later combination of forecasts for the same outcomes. *International Journal of Forecasting*, *20*(3), 391–409. https://doi.org/10.1016/j.ijforecast.2003.09.012

Harvey, N., Harries, C., & Fischer, I. (2000). Using Advice and Assessing Its Quality. *Organizational Behavior and Human Decision Processes*, *81*(2), 252–273. https://doi.org/10.1006/obhd.1999.2874

Heath, C., & Gonzalez, R. (1995). Interaction with Others Increases Decision Confidence but Not Decision Quality: Evidence against Information Collection Views of Interactive Decision Making. *Organizational Behavior and Human Decision Processes*, *61*(3), 305–326. https://doi.org/10.1006/obhd.1995.1024

Hedlund, J., Ilgen, D. R., & Hollenbeck, J. R. (1998). Decision Accuracy in Computer-Mediated versus Face-to-Face Decision-Making Teams. *Organizational Behavior and Human Decision Processes*, *76*(1), 30–47. https://doi.org/10.1006/obhd.1998.2796

Herzog, S. M., & Hertwig, R. (2009). The Wisdom of Many in One Mind: Improving Individual Judgments With Dialectical Bootstrapping. *Psychological Science*, *20*(2), 231–237. https://doi.org/10.1111/j.1467-9280.2009.02271.x

Hollenbeck, J. R., Ilgen, D. R., LePine, J. A., Colquitt, J. A., & Hedlund, J. (1998). Extending the multilevel theory of team decision making: Effects of feedback and experience in hierarcical teams. *Academy of Management Journal*, *41*(3), 269–282. https://doi.org/10.2307/256907

Hollenbeck, J. R., Ilgen, D. R., Sego, D. J., Hedlund, J., Major, D. A., & Phillips, J. (1995). Multilevel theory of team decision making: Decision performance in teams incorporating distributed expertise. *Journal of Applied Psychology*, *80*(2), 292–316. https://doi.org/10.1037/0021-9010.80.2.292

Horowitz, L. M., Krasnoperova, E. N., Tatar, D. G., Hansen, M. B., Person, E. A., Galvin, K. L., & Nelson, K. L. (2001). The Way to Console May Depend on the Goal: Experimental Studies of Social Support. *Journal of Experimental Social Psychology*, *37*(1), 49–61. https://doi.org/10.1006/jesp.2000.1435

Humphrey, S. E., Hollenbeck, J. R., Meyer, C. J., & Ilgen, D. R. (2002). Hierarchical team decision making. In *Research in Personnel and Human Resources Management* (Vol. 21, pp. 175–213). Emerald (MCB UP ). https://doi.org/10.1016/S0742-7301(02)21004-X

Johnson, T. R., Budescu, D. V., & Wallsten, T. S. (2001). Averaging probability judgments: Monte Carlo analyses of asymptotic diagnostic value. *Journal of Behavioral Decision Making*, *14*(2), 123–140. https://doi.org/10.1002/bdm.369

Kennedy, J., Kleinmuntz, D. N., & Peecher, M. E. (1997). Determinants of the Justifiability of Performance in Ill-Structured Audit Tasks. *Journal of Accounting Research*, *35*, 105. https://doi.org/10.2307/2491456

Klayman, J., Soll, J. B., González-Vallejo, C., & Barlas, S. (1999). Overconfidence: It Depends on How, What, and Whom You Ask. *Organizational Behavior and Human Decision Processes*, *79*(3), 216–247. https://doi.org/10.1006/obhd.1999.2847

Kroll, A. (2013). The Other Type of Performance Information: Nonroutine Feedback, Its Relevance and Use. *Public Administration Review*, *73*(2), 265–276. https://doi.org/10.1111/j.1540-6210.2012.02648.x

Krueger, J. I. (2003). Return of the ego--Self-referent information as a filter for social prediction: Comment on Karniol (2003). *Psychological Review*, *110*(3), 585–590. https://doi.org/10.1037/0033-295X.110.3.585

Kruger, J. (1999). Lake Wobegon be gone! The "below-average effect" and the egocentric nature of comparative ability judgments. *Journal of Personality and Social Psychology*, *77*(2), 221–232. https://doi.org/10.1037/0022-3514.77.2.221

Larrick, R. P., Burson, K. A., & Soll, J. B. (2007). Social comparison and confidence: When thinking you're better than average predicts overconfidence (and when it does not). *Organizational Behavior and Human Decision Processes*, *102*(1), 76–94. https://doi.org/10.1016/j.obhdp.2006.10.002

Larrick, R. P., & Soll, J. B. (2006). Intuitions About Combining Opinions: Misappreciation of the Averaging Principle. *Management Science*, *52*(1), 111–127. https://doi.org/10.1287/mnsc.1050.0459

LePine, J. A., Hollenbeck, J. R., Ilgen, D. R., & Hedlund, J. (1997). Effects of individual differences on the performance of hierarchical decision-making teams: Much more than g. *Journal of Applied Psychology*, *82*(5), 803–811. https://doi.org/10.1037/0021-9010.82.5.803

Libby, R., & Blashfield, R. K. (1978). Performance of a composite as a function of the number of judges. *Organizational Behavior and Human Performance*, *21*(2), 121–129. https://doi.org/10.1016/0030-5073(78)90044-2

Lim, J. S., & O'Connor, M. (1995). Judgemental adjustment of initial forecasts: Its effectiveness and biases. *Journal of Behavioral Decision Making*, *8*(3), 149–

168. https://doi.org/10.1002/bdm.3960080302

Magnus, J. R., & Peresetsky, A. A. (2018). Grade Expectations: Rationality and Overconfidence. *Frontiers in Psychology*, *8*, 2346. https://doi.org/10.3389/fpsyg.2017.02346

Maister, D. H., Green, C. H., & Galford, R. M. (2001). *The Trusted Advisor*. NY: Touchstone.

Mannes, A. E. (2009). Are We Wise About the Wisdom of Crowds? The Use of Group Judgments in Belief Revision. *Management Science*, *55*(8), 1267–1279. https://doi.org/10.1287/mnsc.1090.1031

Mannes, A. E., & Moore, D. A. (2013). A Behavioral Demonstration of Overconfidence in Judgment. *Psychological Science*, *24*(7), 1190–1197. https://doi.org/10.1177/0956797612470700

Marteau, T. M., Johnston, M., Wynne, G., & Evans, T. R. (1989). Cognitive factors in the explanation of the mismatch between confidence and competence in performing basic life support. *Psychology & Health*, *3*(3), 173–182. https://doi.org/10.1080/08870448908400377

Mathieu, J., Maynard, M. T., Rapp, T., & Gilson, L. (2008). Team Effectiveness 1997-2007: A Review of Recent Advancements and a Glimpse Into the Future. *Journal of Management*, *34*(3), 410–476. https://doi.org/10.1177/0149206308316061

Meshi, D., Biele, G., Korn, C. W., & Heekeren, H. R. (2012). How Expert Advice Influences Decision Making. *PLoS ONE*, *7*(11), e49748. https://doi.org/10.1371/journal.pone.0049748

Micklethwait, J., & Wooldridge, A. (1996). *The witch doctors: Making sense of the management gurus*. New York: Time Books.

Min, B. H., Oh, Y., & Brower, R. S. (2020). The Effects of Diverse Feedback Dynamics on Performance Improvement: A Typology of Performance Feedback Signals. *Administration & Society*, 009539972093382. https://doi.org/10.1177/0095399720933822

Minson, J. A., Liberman, V., & Ross, L. (2011). Two to Tango: Effects of Collaboration and Disagreement on Dyadic Judgment. *Personality and Social Psychology Bulletin*, *37*(10), 1325–1338. https://doi.org/10.1177/0146167211410436

Minson, J. A., & Mueller, J. S. (2012). The Cost of Collaboration: Why Joint Decision Making Exacerbates Rejection of Outside Information. *Psychological Science*, *23*(3), 219–224. https://doi.org/10.1177/0956797611429132

Moore, D. A., Carter, A. B., & Yang, H. H. J. (2015). Wide of the mark: Evidence on the underlying causes of overprecision in judgment. *Organizational Behavior and Human Decision Processes*, *131*, 110–120. https://doi.org/10.1016/j.obhdp.2015.09.003

Moore, D. A., & Dev, A. S. (2018). Individual differences in overconfidence. *Encyclopedia of Personality and Individual Differences*, 1–5. https://doi.org/10.1037/pspa0000018

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review*, *115*(2), 502–517. https://doi.org/10.1037/0033-295X.115.2.502

Moore, D. A., & Kim, T. G. (2003). Myopic Social Prediction and the Solo Comparison Effect. *Journal of Personality and Social Psychology*, *85*(6), 1121–1135. https://doi.org/10.1037/0022-3514.85.6.1121

Moore, D. A., & Schatz, D. (2017). The three faces of overconfidence. *Social and Personality Psychology Compass*, *11*(8), e12331. https://doi.org/10.1111/spc3.12331

Moore, D. A., & Small, D. A. (2007). Error and bias in comparative judgment: On being both better and worse than we think we are. *Journal of Personality and Social Psychology*, *92*(6), 972–989. https://doi.org/10.1037/0022-3514.92.6.972

Moynihan, D. P., & Pandey, S. K. (2010). The Big Question for Performance Management:

Why Do Managers Use Performance Information? *Journal of Public Administration Research and Theory*, *20*(4), 849–866. https://doi.org/10.1093/jopart/muq004

Moynihan, Donald P., & Lavertu, S. (2012). Does Involvement in Performance Management Routines Encourage Performance Information Use? Evaluating GPRA and PART. *Public Administration Review*, *72*(4), 592–602. https://doi.org/10.1111/j.1540-6210.2011.02539.x

Neale, M. A., & Bazerman, M. H. (1991). *Cognition and Rationality in Negotiation*. https://www.hbs.edu/faculty/Pages/item.aspx?num=5084

Otten, W., & Van Der Pligt, J. (1996). Context Effects in the Measurement of Comparative Optimism in Probability Judgments. *Journal of Social and Clinical Psychology*, *15*(1), 80–101. https://doi.org/10.1521/jscp.1996.15.1.80

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(3), 534–552. https://doi.org/10.1037/0278-7393.14.3.534

Pedregon, C. A., Farley, R. L., Davis, A., Wood, J. M., & Clark, R. D. (2012). Social desirability, personality questionnaires, and the "better than average" effect. *Personality and Individual Differences*, *52*(2), 213–217. https://doi.org/10.1016/j.paid.2011.10.022

Perloff, L. S., & Fetzer, B. K. (1986). Self–other judgments and perceived vulnerability to victimization. *Journal of Personality and Social Psychology*, *50*(3), 502–510. https://doi.org/10.1037/0022-3514.50.3.502

Phillips, J. M. (1999). Antecedents of Leader Utilization of Staff Input in Decision-Making Teams. *Organizational Behavior and Human Decision Processes*, *77*(3), 215–242. https://doi.org/10.1006/obhd.1998.2819

Price, P. C., & Stone, E. R. (2004). Intuitive evaluation of likelihood judgment producers: Evidence for a confidence heuristic. *Journal of Behavioral Decision Making*, *17*(1), 39–57. https://doi.org/10.1002/bdm.460

Proeger, T., & Meub, L. (2014). Overconfidence as a social bias: Experimental evidence. *Economics Letters*, *122*(2), 203–207.

Rader, C. A., Soll, J. B., & Larrick, R. P. (2015). Pushing away from representative advice: Advice taking, anchoring, and adjustment. *Organizational Behavior and Human Decision Processes*, *130*, 26–43. https://doi.org/10.1016/j.obhdp.2015.05.004

Redd, S. B. (2002). The Influence of Advisers on Foreign Policy Decision Making: An Experimental Study. *Journal of Conflict Resolution*, *46*(3), 335–364. https://doi.org/10.1177/0022002702046003002

Ross, L., Lepper, M. R., & Hubbard, M. (1975). Perseverance in self-perception and social perception: Biased attributional processes in the debriefing paradigm. *Journal of Personality and Social Psychology*, *32*(5), 880–892. https://doi.org/10.1037/0022-3514.32.5.880

Savadori, L., Van Swol, L. M., & Sniezek, J. A. (2001). Information Sampling and Confidence Within Groups and Judge Advisor Systems. *Communication Research*, *28*(6), 737–771. https://doi.org/10.1177/009365001028006002

Schaefer, P. S., Williams, C. C., Goodie, A. S., & Campbell, W. K. (2004). Overconfidence and the Big Five. *Journal of Research in Personality*, *38*(5), 473–480.

Schotter, A. (2003). Decision Making with Naive Advice. *EXPERIMENTAL ECONOMICS*, *93*(2), 196–201.

Schrah, G. E., Dalal, R. S., & Sniezek, J. A. (2006). No decision-maker is an Island: Integrating expert advice with information acquisition. *Journal of Behavioral Decision Making*, *19*(1), 43–60. https://doi.org/10.1002/bdm.514

Schultze, T., Mojzisch, A., & Schulz-Hardt, S. (2013). Groups Weight Outside Information Less Than Individuals Do Because They Should: Response to Minson and Mueller (2012). *Psychological Science*, *24*(7), 1371–1372.

https://doi.org/10.1177/0956797612472206

Schultze, T., Mojzisch, A., & Schulz-Hardt, S. (2017). On the inability to ignore useless advice: A case for anchoring in the judge-advisor-system. *Experimental Psychology*, *64*(3), 170–183. https://doi.org/10.1027/1618-3169/a000361

See, K. E., Morrison, E. W., Rothman, N. B., & Soll, J. B. (2011). The detrimental effects of power on confidence, advice taking, and accuracy. *Organizational Behavior and Human Decision Processes*, *116*(2), 272–285. https://doi.org/10.1016/j.obhdp.2011.07.006

Sniezek, J. A., & Buckley, T. (1995). Cueing and Cognitive Conflict in Judge-Advisor Decision Making. *Organizational Behavior and Human Decision Processes*, *62*(2), 159–174. https://doi.org/10.1006/obhd.1995.1040

Sniezek, J. A., & Henry, R. A. (1989). Accuracy and confidence in group judgment. *Organizational Behavior and Human Decision Processes*, *43*(1), 1–28. https://doi.org/10.1016/0749-5978(89)90055-1

Sniezek, J. A., Schrah, G. E., & Dalal, R. S. (2004). Improving judgement with prepaid expert advice. *Journal of Behavioral Decision Making*, *17*(3), 173–190. https://doi.org/10.1002/bdm.468

Sniezek, J. A., & Van Swol, L. M. (2001). Trust, Confidence, and Expertise in a Judge-Advisor System. *Organizational Behavior and Human Decision Processes*, *84*(2), 288–307. https://doi.org/10.1006/obhd.2000.2926

Soll, J. B., & Klayman, J. (2004). Overconfidence in Interval Estimates. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 299–314. https://doi.org/10.1037/0278-7393.30.2.299

Soll, J. B., & Larrick, R. P. (2009). Strategies for revising judgment: How (and how well) people use others' opinions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(3), 780–805. https://doi.org/10.1037/a0015145

Soll, J. B., & Mannes, A. E. (2011). Judgmental aggregation strategies depend on whether the self is involved. *International Journal of Forecasting*, *27*(1), 81–102. https://doi.org/10.1016/j.ijforecast.2010.05.003

Sorkin, R. D., Hays, C. J., & West, R. (2001). Signal-detection analysis of group decision making. *Psychological Review*, *108*(1), 183–203. https://doi.org/10.1037/0033-295X.108.1.183

Stavrova, O., Köneke, V., & Schlösser, T. (2016). Overfulfilling the norm: The better-than-average effect in judgments of attitudes. *Social Psychology*, *47*(5), 288–293. https://doi.org/10.1027/1864-9335/a000280

Tost, L. P., Gino, F., & Larrick, R. P. (2012). Power, competitiveness, and advice taking: Why the powerful don't listen. *Organizational Behavior and Human Decision Processes*, *117*(1), 53–65. https://doi.org/10.1016/j.obhdp.2011.10.001

Tversky, A., & Kahneman, D. (1974). *Judgment under Uncertainty: Heuristics and Biases*. *185*, 10.

Tzioti, S. C., Wierenga, B., & van Osselaer, S. M. J. (2014). The Effect of Intuitive Advice Justification on Advice Taking: Effect of Intuitive Advice Justification. *Journal of Behavioral Decision Making*, *27*(1), 66–77. https://doi.org/10.1002/bdm.1790

Van Swol, L. M. (2011). Forecasting another's enjoyment versus giving the right answer: Trust, shared values, task effects, and confidence in improving the acceptance of advice. *International Journal of Forecasting*, *27*(1), 103–120. https://doi.org/10.1016/j.ijforecast.2010.03.002

Van Swol, L. M., Paik, J. E., & Prahl, A. (2018). The Oxford Handbook of Advice. In E. L. MacGeorge & L. M. Van Swol (Eds.), *Advice Recipients: The Psychology of Advice Utilization* (Vol. 1). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190630188.013.2

Van Swol, L. M., & Sniezek, J. A. (2005). Factors affecting the acceptance of expert advice. *British Journal of Social Psychology*,

*44*(3), 443–461. https://doi.org/10.1348/014466604X17092

Weinstein, N. D. (1980). Unrealistic optimism about future life events. *Journal of Personality and Social Psychology*, *39*(5), 806–820. https://doi.org/10.1037/0022-3514.39.5.806

Whittemore, R., Rankin, S. H., Callahan, C. D., Leder, M. C., & Carroll, D. L. (2000). The Peer Advisor Experience Providing Social Support. *Qualitative Health Research*, *10*(2), 260–276. https://doi.org/10.1177/104973200129118408

Windschitl, P. D., Kruger, J., & Simms, E. N. (2003). The Influence of Egocentrism and Focalism on People's Optimism in Competitions: When What Affects Us Equally Affects Me More. *Journal of Personality and Social Psychology*, *85*(3), 389–408. https://doi.org/10.1037/0022-3514.85.3.389

Winkler, R. L., & Poses, R. M. (1993). Evaluating and Combining Physicians' Probabilities of Survival in an Intensive Care Unit. *Management Science*, *39*(12), 1526–1543. https://doi.org/10.1287/mnsc.39.12.1526

Yaniv, I. (1997). Weighting and Trimming: Heuristics for Aggregating Judgments under Uncertainty. *Organizational Behavior and Human Decision Processes*, *69*(3), 237–249. https://doi.org/10.1006/obhd.1997.2685

Yaniv, I. (2004a). Receiving other people's advice: Influence and benefit. *Organizational Behavior and Human Decision Processes*, *93*(1), 1–13. https://doi.org/10.1016/j.obhdp.2003.08.002

Yaniv, I. (2004b). The Benefit of Additional Opinions. *Current Directions in Psychological Science*, *13*(2), 75–78. https://doi.org/10.1111/j.0963-7214.2004.00278.x

Yaniv, I., & Choshen-Hillel, S. (2012). Exploiting the Wisdom of Others to Make Better Decisions: Suspending Judgment Reduces Egocentrism and Increases Accuracy: Exploiting the Wisdom of Others. *Journal of Behavioral Decision Making*, *25*(5), 427–434. https://doi.org/10.1002/bdm.740

Yaniv, I., Choshen-Hillel, S., & Milyavsky, M. (2011). Receiving advice on matters of taste: Similarity, majority influence, and taste discrimination. *Organizational Behavior and Human Decision Processes*, *115*(1), 111–120. https://doi.org/10.1016/j.obhdp.2010.11.006

Yaniv, I., & Hogarth, R. M. (1993). Judgmental Versus Statistical Prediction: Information Asymmetry and Combination Rules. *Psychological Science*, *4*(1), 58–62. https://doi.org/10.1111/j.1467-9280.1993.tb00558.x

Yaniv, I., & Kleinberger, E. (2000). Advice Taking in Decision Making: Egocentric Discounting and Reputation Formation. *Organizational Behavior and Human Decision Processes*, *83*(2), 260–281. https://doi.org/10.1006/obhd.2000.2909

Yaniv, I., & Milyavsky, M. (2007). Using advice from multiple sources to revise and improve judgments. *Organizational Behavior and Human Decision Processes*, *103*(1), 104–120. https://doi.org/10.1016/j.obhdp.2006.05.006

Yates, J. F., Price, P. C., Lee, J.-W., & Ramirez, J. (1996). Good probabilistic forecasters: The 'consumer's' perspective. *International Journal of Forecasting*, *12*(1), 41–56. https://doi.org/10.1016/0169-2070(95)00636-2

Zakay, D., & Glicksohn, J. (1992). Overconfidence in a Multiple-Choice Test and its Relationship to Achievement. *The Psychological Record*, *42*(4), 519–524. https://doi.org/10.1007/BF03399619

Zarnoth, P., & Sniezek, J. A. (1997). The Social Influence of Confidence in Group Decision Making. *Journal of Experimental Social Psychology*, *33*(4), 345–366. https://doi.org/10.1006/jesp.1997.1326

Zarnowitz, V. (1984). The accuracy of individual and group forecasts from business outlook surveys. *Journal of Forecasting*, *3*(1), 11–26. https://doi.org/10.1002/for.3980030103

Zell, E., & Alicke, M. D. (2011). Age and the Better-Than-Average Effect: BETTER-THAN-AVERAGE EFFECT. *Journal of*

*Applied Social Psychology*, *41*(5), 1175–1188. https://doi.org/10.1111/j.1559-1816.2011.00752.x

Zhou, Q., Yin, C., & Liu, Z. (2020). The Effect of Time Pressure and Task Difficulty on Human Search. In D. Harris & W.-C. Li (Eds.), *Engineering Psychology and Cognitive Ergonomics. Mental Workload, Human Physiology, and Human Energy: 17th International Conference, EPCE 2020, Held as Part of the 22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings, Part I* (Vol. 12186). Springer International Publishing. https://doi.org/10.1007/978-3-030-49044-7

## Appendix

### List of Tables



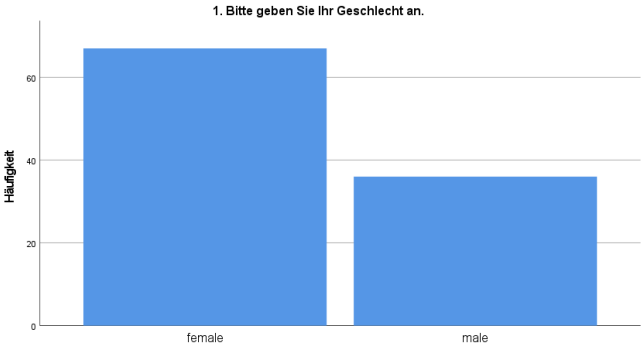
### List of Figures



### List of Appendices

**1. Bitte geben Sie Ihr Geschlecht an.**

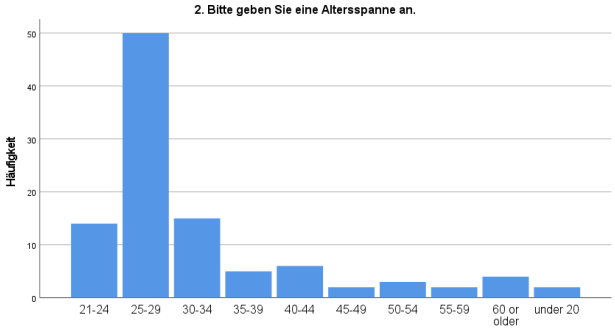
| | | Häufigkeit | Prozent | Gültige Prozente | Kumulierte Prozente |
|---|---|---|---|---|---|
| Gültig | female | 67 | 65,0 | 65,0 | 65,0 |
| | male | 36 | 35,0 | 35,0 | 100,0 |
| | Gesamt | 103 | 100,0 | 100,0 | |




*Appendix 1: Distribution of Gender*

**2. Bitte geben Sie eine Altersspanne an.**

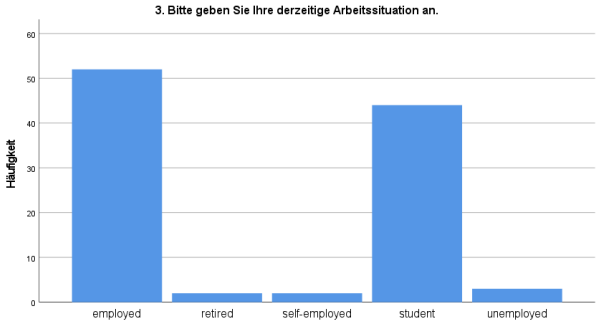
| | | Häufigkeit | Prozent | Gültige Prozente | Kumulierte Prozente |
|---|---|---|---|---|---|
| Gültig | 21-24 | 14 | 13,6 | 13,6 | 13,6 |
| | 25-29 | 50 | 48,5 | 48,5 | 62,1 |
| | 30-34 | 15 | 14,6 | 14,6 | 76,7 |
| | 35-39 | 5 | 4,9 | 4,9 | 81,6 |
| | 40-44 | 6 | 5,8 | 5,8 | 87,4 |
| | 45-49 | 2 | 1,9 | 1,9 | 89,3 |
| | 50-54 | 3 | 2,9 | 2,9 | 92,2 |
| | 55-59 | 2 | 1,9 | 1,9 | 94,2 |
| | 60 or older | 4 | 3,9 | 3,9 | 98,1 |
| | under 20 | 2 | 1,9 | 1,9 | 100,0 |
| | Gesamt | 103 | 100,0 | 100,0 | |




*Appendix 2: Distribution of Age*

**3. Bitte geben Sie Ihre derzeitige Arbeitssituation an.**

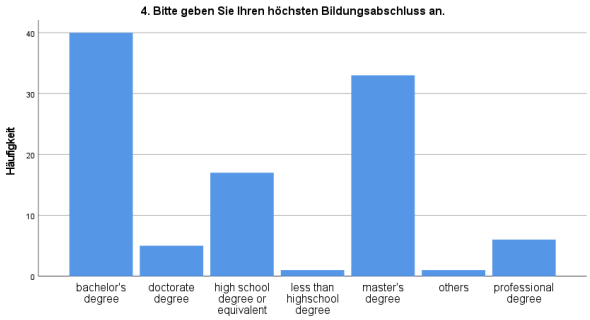
| | | Häufigkeit | Prozent | Gültige Prozente | Kumulierte Prozente |
|---|---|---|---|---|---|
| Gültig | employed | 52 | 50,5 | 50,5 | 50,5 |
| | retired | 2 | 1,9 | 1,9 | 52,4 |
| | self-employed | 2 | 1,9 | 1,9 | 54,4 |
| | student | 44 | 42,7 | 42,7 | 97,1 |
| | unemployed | 3 | 2,9 | 2,9 | 100,0 |
| | Gesamt | 103 | 100,0 | 100,0 | |




*Appendix 3: Distribution of Employment Status*

**4. Bitte geben Sie Ihren höchsten Bildungsabschluss an.**

| | | Häufigkeit | Prozent | Gültige Prozente | Kumulierte Prozente |
|---|---|---|---|---|---|
| Gültig | bachelor's degree | 40 | 38,8 | 38,8 | 38,8 |
| | doctorate degree | 5 | 4,9 | 4,9 | 43,7 |
| | high school degree or equivalent | 17 | 16,5 | 16,5 | 60,2 |
| | less than highschool degree | 1 | 1,0 | 1,0 | 61,2 |
| | master's degree | 33 | 32,0 | 32,0 | 93,2 |
| | others | 1 | 1,0 | 1,0 | 94,2 |
| | professional degree | 6 | 5,8 | 5,8 | 100,0 |
| | Gesamt | 103 | 100,0 | 100,0 | |




*Appendix 4: Distribution of Education Status*

**89. Denken Sie in Ihrem Alltag an die Kilokalorienmenge von Gerichten, die Sie essen?**

| | | Häufigkeit | Prozent | Gültige Prozente | Kumulierte Prozente |
|---|---|---|---|---|---|
| Gültig | no | 58 | 56,3 | 56,3 | 56,3 |
| | yes | 45 | 43,7 | 43,7 | 100,0 |
| | Gesamt | 103 | 100,0 | 100,0 | |

**90. Spielt die Kilokalorienmenge von Gerichten in Ihrem Alltag eine wichtige Rolle?**

| | | Häufigkeit | Prozent | Gültige Prozente | Kumulierte Prozente |
|---|---|---|---|---|---|
| Gültig | no | 74 | 71,8 | 71,8 | 71,8 |
| | yes | 29 | 28,2 | 28,2 | 100,0 |
| | Gesamt | 103 | 100,0 | 100,0 | |

*Appendix 5: Distribution of Attitude towards Kilocalories*

**Umfragesprache**

| | | Häufigkeit | Prozent | Gültige Prozente | Kumulierte Prozente |
|---|---|---|---|---|---|
| Gültig | DE | 54 | 52,4 | 52,4 | 52,4 |
| | EN | 49 | 47,6 | 47,6 | 100,0 |
| | Gesamt | 103 | 100,0 | 100,0 | |

*Appendix 6: Distribution of Survey Language*

**Tests auf Univariate**

| Quelle | Maß | | Quadratsumme vom Typ III | df | Mittel der Quadrate | F | Sig. | Partielles Eta-Quadrat |
|---|---|---|---|---|---|---|---|---|
| Time * Level_ifficulty | WOA | Sphärizität angenommen | ,418 | 5 | ,084 | 1,339 | ,246 | ,013 |
| | | Greenhouse-Geisser | ,418 | 2,842 | ,147 | 1,339 | ,263 | ,013 |
| | | Huynh-Feldt | ,418 | 2,935 | ,142 | 1,339 | ,262 | ,013 |
| | | Untergrenze | ,418 | 1,000 | ,418 | 1,339 | ,250 | ,013 |
| | JudgAcc | Sphärizität angenommen | 4635959,740 | 5 | 927191,948 | 42,416 | ,000 | ,300 |
| | | Greenhouse-Geisser | 4635959,740 | 3,244 | 1428961,535 | 42,416 | ,000 | ,300 |
| | | Huynh-Feldt | 4635959,740 | 3,367 | 1376799,860 | 42,416 | ,000 | ,300 |
| | | Untergrenze | 4635959,740 | 1,000 | 4635959,740 | 42,416 | ,000 | ,300 |
| Time * Level_Difficulty | Overestimation | Sphärizität angenommen | 2608572,497 | 5 | 521714,499 | 27,540 | ,000 | ,218 |
| | | Greenhouse-Geisser | 2608572,497 | 3,992 | 653421,546 | 27,540 | ,000 | ,218 |
| | | Huynh-Feldt | 2608572,497 | 4,181 | 623927,358 | 27,540 | ,000 | ,218 |
| | | Untergrenze | 2608572,497 | 1,000 | 2608572,497 | 27,540 | ,000 | ,218 |
| | Overprecision | Sphärizität angenommen | 9,317 | 5 | 1,863 | 10,060 | ,000 | ,092 |
| | | Greenhouse-Geisser | 9,317 | 4,875 | 1,911 | 10,060 | ,000 | ,092 |
| | | Huynh-Feldt | 9,317 | 5,000 | 1,863 | 10,060 | ,000 | ,092 |
| | | Untergrenze | 9,317 | 1,000 | 9,317 | 10,060 | ,002 | ,092 |
| | Overplacement | Sphärizität angenommen | 15,217 | 5 | 3,043 | ,284 | ,922 | ,003 |
| | | Greenhouse-Geisser | 15,217 | 4,571 | 3,329 | ,284 | ,909 | ,003 |
| | | Huynh-Feldt | 15,217 | 4,819 | 3,157 | ,284 | ,917 | ,003 |
| | | Untergrenze | 15,217 | 1,000 | 15,217 | ,284 | ,596 | ,003 |

*Appendix 7: Analysis on Variables*

**Abstract**

Die nachfolgende Arbeit betrachtet den Einfluss von Aufgabenschwierigkeit und Leistungsrückmeldung auf Ratschlagsannahme, Urteilspräzision und übersteigertes Selbstvertrauen während des Ratschlags-Annahme-Prozesses. Ziel dieser Arbeit ist die Beantwortung der Forschungsfrage, ob Teilnehmer lernen, Ratschläge im Zeitverlauf optimal zu nutzen, wenn sie Leistungsrückmeldung erhalten. Drei Hypothesen wurden aufgestellt, um die Entwicklung der Ratschlagsannahme, Urteilspräzision und den drei Formen von übersteigertem Selbstvertrauen zu testen. 103 Probanden gaben Schätzungen bezüglich der Kilokalorienanzahl von Gerichten sowie Information zu ihrem Selbstvertrauen an. Die Ergebnisse zeigten, dass die Ratschlagsannahme niedrig war und Teilnehmer eher Zeichen von Unterschätzung anstatt von Überschätzung zeigten. Dennoch nahm die Urteilspräzision über die Runden hinweg zu, da die Teilnehmer lernten, präzisere Entscheidungen zu treffen und ihr 90% Konfidenzintervall anzupassen, was zu einer geringeren Überpräzision führte. Bezüglich der Überplatzierung konnte keine Aussage gemacht werden, da die Ergebnisse über die Runden zu sehr Schwankungen unterlagen.