DISSERTATION

Titel der Dissertation

"Role of mutual identification in motivating employee's labor donations within employee-customer dyad"

Verfasserin

Oksana Galak

angestrebter akademischer Grad

Doctor of Philosophy (PhD)

Wien, 2015

Studienkennzahl lt. Studienblatt:	A 094 151 146
Dissertationsgebiet lt. Studienblatt:	PhD-Studium (Doctor of Philosophy) UG2002 (Dissertationsgebiet: Betriebswirtschaft) Management
Betreuer:	o. Univ.-Prof. Mag. Dr. Rudolf Vetschera

Attestation

I certify that this Dissertation reports my original work accomplished during my study in the PhD program in Management at the University of Vienna (Austria).

Acknowledgements

Table of Contents

List of Figures

List of Tables

dependent variable

different parameters of output (group split)

List of Appendixes

Motivate - provide (someone) with a reason for doing something.

Oxford English Dictionary

I.   Introduction

Job motivation is a complex issue discussed in multiple fields of organizational sciences from different angles. One of the definitions which can serve a good starting point for our discussion characterizes job motivation as the will to achieve the best performance an employee is able to generate based on her strengths and competences in order to reach company's goals (Calabrese, 2012). What kind of mechanisms can be employed in order to motivate employees to act in order to accomplish company's goals?

The answer to this question in the literature depends on the views of the researchers on the human nature. The two most extreme streams of thought see people either as purely self-interested, or as benevolent. In the first case the right incentives (usually monetary) should be used to align the interests of the employee with organizational goals, in the second case good behavior is seen as natural and a bad one - as a result of wrong premises about people and wrong incentives (Ben-Ner, 2013).

In the standard economic models the first view prevails, and the main tool for motivating employees is to use incentives which are linking the agent's compensation to her performance (Edele et al, 2013). In order to do that efficiently management should be aware of how the relationship between the employees' effort and her performance precisely looks like.

When output is perfectly observable, its value provides a perfect indicator of agent's effort and it is relatively easy to incentivize optimal level of effort. However, when effort also depends on some random component, it is more difficult to determine the agent's effort precisely. Moreover in many of the modern occupations agents perform different tasks in their jobs and multitasking complicates the optimal design of incentive scheme (Holmstrom and Milgrom, 1987; Simon, 1991). When different attributes of the task compete for resources the motivational tools should influence not only the general amount of effort, but also the trade-off among attributes.

There are many ways in which job performance can be operationalized, with two dimensions consistently emerging from the research: quality and quantity, which are even identified as main components of performance (Graso and Probst, 2012). Under certain conditions quantity and quality compete for resources, which complicates the ways of measuring and interpreting the output.

Complex structures of output requiring trade-offs between different tasks or between various attributes of one task are common for modern-world jobs. Being reflected in the incompleteness of contractual arrangements between employer and employee, this situation leads to the concentration of the decision power regarding the distribution of the effort in the hands of the employee. As Simon (1991, p.32)

puts it "doing the job well is not mainly the matter of responding to commands, but is much more a matter of taking initiative to advance organizational objectives".

In complex, multidimensional tasks the effort of employees, we believe, needs to be guided and supported by a wider array of motivational mechanisms than is used in tasks with easier measurable outputs.

There are fields where this need for multifaceted motivational mechanism is especially strong. For example the biggest challenge for jobs in the service sector comes from the necessity to increase both service productivity and service quality. According to Calabrese (2012) technical, organizational, or managerial resources by themselves can increase one but not both parameters. To overcome the productivity-quality trade-off the researchers propose to focus on human resources and on leveraging their motivations.

Also for public sector organizations which lack the single, clear goal of value maximization of private sector, the challenge of incentivizing employees (as well as measuring the output) is hard (Burgess and Ratto, 2003).

In general, when we talk about a complex construct of customer satisfaction (or client well-being) as a goal of organisation, then the need for the motivational tools able to incentivise not only the general amount of effort, but also its optimal distribution (given the level of task proficiency of employees) becomes especially challenging.

While classical models of motivation were mostly built on the premises that preferences are driven by money, a wide array of non-pecuniary tastes (preferences), especially in the form of social preferences, was included in the past 20 years into the utility function (Huettel and Kranton, 2012).[1]

These non-pecuniary preferences, like altruism or fairness, are seen in the theory as generic: some people are more altruistic or fair than the others. For example, the general observation of the research in the field of public economics is that non-profit organizations depend on such preferences (Prendergast, 2007, 2008, Francois and Vlassopoulos, 2008). Non-profit workers stress that the impact their work makes on others provides them with motivation, which goes beyond monetary rewards. The argument is that due to their specific type of goals these organizations are able to attract prosocial (benevolent) employees (or such employees self-select into these organizations).

We argue that not only this type of organisations can benefit from prosocial behavior of their employees. Indeed, we should take into account that in most of the organizations employees are heterogeneous in respect to their preferences, which is why there is a need to build the structure of incentives which takes this heterogeneity into account, not limiting itself to the pure benevolence or self-interest (Ben-Ner, 2013).

---

[1] Literature in organizational behavior also acknowledges differences in individual motivations, but "has not put forth a comprehensive view of human nature" (Ben-Ner, 2013).

We adopt a view that social preferences, being an important motivator, depend in large part on people's identities and social norms for behaviour in different contexts (Huettel and Kranton, 2012). For the purposes of our discussion this means that organizations can motivate their employees to behave prosocially, also towards their customers, if they create the contexts where social preferences come into light.

Given the nature of the output, managers and workers may care about the output, which is especially relevant for the jobs in the public sector (Burgess and Ratto, 2003). The important premise is whether the welfare of clients is the goal of organization itself. In this case workers may internalize this goal and exert effort on behalf of the client.[2]

There are multiple underlying dimensions which can be employed by the organisations in order to create the context where employees are motivated to place more value on the utility of their customers. We see our general goal in defining the contextual factors able to promote prosocial behavior in employee-customer dyad (in other words, pro-customer behavior) and further testing the strengths of such factors empirically.

In our theorizing we widely use the result of research in different fields, with the major emphasis on the experimental economics, which nowadays is very active in informing science with new insights into the specifics of other-regarding behavior.

While it is true that experimental economics provides much of support for the other-regarding preferences, its application to a wider field of employee motivation (beyond the issue of self-selection in particular types of organisations) requires context-dependent approach. The researches in this area admit that people can be altruistic towards relatives and strangers, but seemingly not to those who finally benefit from the results of their work effort. In our view customer-oriented behavior in form of the effort provided by the employees on behalf of the customer without pecuniary incentives is a promising direction of managerial application of the idea of social preferences.

We open our discussion with the theoretical overview of the origins of social preferences, starting from classical notion of altruism, which is the first social preference researched within economics. We can see that altruism is a very mixed concept, defined differently in various disciplines and widely discussed even within economic domain. Classical altruism was unable to explain many facts about prosocial behavior. However, modern theories of social preferences also face challenges while explaining the roots of such behavior (Povey, 2014). We argue that the reason is in the fact that both in traditional and in new definitions of prosocial behavior are discussed from a trait-like perspective.

That is why in our next step we leave the narrow trait-based frame of prosocial behavior and discuss the possible reasons for behaving in other-regarding way. The first look here is thrown at emotional mechanisms governing behavior and we talk in particular about empathy and emphatic concern and

---

[2] Interestingly in this case introducing financial rewards entirely based on performance may signal to the workers that the relationship between them and the organisation is a pure market one, which reduces worker's prosocial motivation (Burgess and Ratto, 2003).

how they influence the behavior. In our search for contextual factors mediating empathic feelings we discover that almost every research talking about emphatic concern stresses the importance of information available to the counterparties. It is especially true for the amount and content of the information about the recipient / beneficiary available to the giver. Research argues that such information increases the ability of the giver to identify with the recipient and thus, promotes other-regarding behavior.

Presence of information about the other person reduces social distance between this person and oneself.[3] And since people are expected to act more favourably toward those with a higher degree of social kinship, the smaller the social distance the more other-regarding behavior will be shown.

The results can be twofold: increase in the general level of effort and increase in the willingness to satisfy the customer (as utility is gained when the customer is satisfied). Achieving the second goal is closely related to the necessity of finding the right trade-off among several attributes of output, achieving the combination of these attributes best benefiting the customer.

Further we move to another perspective which links other-regarding behavior to the notion of social norms. This research has a long tradition in sociology, but only recently it has been accepted on a wider scale in economics. Norms can be internalized and hence, internally administered. In this case they motivate prosocial behavior through the feeling of "doing the right thing", which is known as warm-glow (Andreoni, 1990). The more important relationship for the purposes of our research, though, is between social norms and observability of the actions. The major idea here is that norms-conforming behavior is strengthened if the actions of the giver can be externally observed. In other words if the information about the giver is communicated to the outer world (e.g. to the recipient or other external parties), the giver exhibits higher prosocial behavior, compared to the cases where the giver is anonymous.

Joining these two perspectives based on the literature used, we come to the conclusion that altering the amount of information available to the parties about each other can play a crucial role in stimulating other-regarding behavior.

We place our research questions into the frame of multiattribute task environment, where various attributes compete for resources (amount of the employee's effort and time). We differentiate between attributes of output having high and low level of measurement ambiguity; having high and low level of importance for customer; being explicitly set in the instructions or implicitly derived by the participants. Importantly the output requirements for all attributes are of maximization nature, without any threshold and monetary punishment.

We hypothesize that providing the employee with the information about the customer leads to her focus on the attributes having higher value to the customer. The underlying reason is the increase in empathic concern and perspective taking prompted by reduction of social distance.

---

[3] Social distance can be understood as the emotional proximity between people induced by the situation (Charness and Gneezy, 2000).

4

At the same time opening the employee's identity to the customer leads to the focus on the attributes easily measurable and signalling the employee's effort. We argue that this happens due to the image concerns activated by increase in observability of employee's effort.

In our experimental design we follow the call of Huettel and Kranton (2012) to incorporate meaningful real-world relationships into experimental paradigms. Researchers claim that in economics social aspects of interaction "are pushed into the background" and "nearly all studies of interpersonal interactions, so far, have used abstract social relationships (e.g. "opponents" or "partners" in games). In reality though few economic interactions are truly anonymous and free of social context and identity. There can be trade-off between different factors influencing the individual (preferences) even in simple situations.

For the testing our hypotheses we employed the experimental method. Our experimental study involved 100 students of the faculty of Business, Economics and Statistics of the University of Vienna.

The major results of the experiment supported the hypotheses and provided some explorative insights in the behavioral patterns of our subjects. Follow-up study, explorative in its nature, was conducted a year after the experiment and involved more than 30 participants of the initial experiment. It supported the interpretation of the results of original experiment and tested a new tool linking the experimental approach with strategy method.

The main results of the research can be summarised as follows:

- Providing employees with information regarding the general purpose of their activities (in our case general information about the customer) leads to the increase in the overall level of effort[4]. The focus of this effort is concentrated on those attributes of the output, which required level is explicitly stated by the management. In the situation where the time constraint requires the trade-off between different attributes an attempt to meet the requirements of the management (without questioning their value for the customer) can lead to the lower focus and effort directed at the attributes having greater value for the customer.

- Providing employees with the information about particular customer (in our case photo and personal information; additionally to the general information) leads to an increase in the overall level of effort with focus on the attributes having greater value for the customer. Beliefs about helping behavior of others play a decisive role in determining trade-offs among different attributes.

- Requiring the employees to reveal their identity to the customer (in our case the request to provide the name to the customer; additionally to the general information) results in the focus towards easily measurable parameters of output, thus, signalling the effort of the employee. In the situation

---

[4] Here and further in this part when we talk about the difference in the effort, we mean in comparison with the no-purpose condition (control time in our experimental setting).

where the time constraint requires the trade-off between different attributes such concentration of attention on the easily measurable parameters of output can lead to the comparatively lower focus on the attributes having greater value for the customer.

- Combining two treatment effects (information about the specific customer and requirement of non-anonymity of the employee; additionally to the general information) leads to a relatively more balanced focus on different attributes of the output with the effect of image concerns slightly overweighting the effect of empathic concerns. Task proficiency plays an essential role in determining the variation among employees.

We conclude by saying that our experimental study was able to show that under conditions of incomplete contracting and fixed remuneration without direct monitoring or financial sanctions slight changes in the identifiability of the customer or / and employee can alter both the amount of effort and the trade-offs between different attributes of the task. Both our analysis of the literature and the design of the experiment itself had a purpose of defining and interpreting the mechanisms behind the context-triggered changes in employee's behavior. However, further empirical studies and observations are necessary to acquire a better understanding of these mechanisms at work.

II.   Literature Review

It is not from the benevolence of the butcher, the brewer, or the baker that we expect our dinner, but from their regard to their own interest.
Adam Smith[5]

To feel much for others and little for ourselves; to restrain our selfishness and Exercise our benevolent affections, constitute the perfection of human nature.
Adam Smith[6]

2.1. General overview

2.1.1.   Classical approach to job motivation and modern challenges

Creative, willing to learn and loyal employees are valuable, often rare and difficult to imitate resource of a company. Since people are not born entirely loyal to their future employer (although different types of personality do have impact on how people behave at work), job motivation - the 'why' of behavior - has been for many years one of the hottest topics of debate among scientists and practitioners from different fields.

According to the Oxford English Dictionary, job is "a paid position of regular employment", and motivation is "a reason or reasons for acting or behaving in a particular way". Putting the two together economists make the emphasis on "paid" and conclude that money is the primary motivation for performing work-related activities.

Following classical economic perspective the employee (as well as any individual) can be seen as rational self-centred profit-maximizer, who is essentially lazy (prefers less effort to more effort) and is concerned only with her own monetary pay-off. In other words the effort exerted to perform the job always has a negative utility for the employee and money always has a positive utility (in the sense that more money is always preferred to less); or if we put it differently, employee is driven by the desire to get as much money as she can while providing as little effort as possible.

Hence, the traditional approach to motivating employees involves creation of optimal incentive schemes based on monetary punishments and rewards in order to assure the optimal level of trade-off between the effort provided by the employee and financial rewards she is expected to get. This tactic relies on the above mentioned assumption of increasing utility of money and disutility of effort.

The classical motivational mechanisms based on the above understanding of human nature used to be quite efficient in the past, but with the time their efficiency is getting increasingly questioned. There are several reasons for this trend. Firstly, the changes in environment for organizations and their

---

[5] An Inquiry Into the Nature and Causes of the Wealth of Nations (1776), Vol. 1, Ch. 2.
[6] The Theory of Moral Sentiments (1759), Pt i, Ch i.

employees have challenged the assumptions about the nature of jobs and performance requirements for employees. In particular, modern organizations often function in rapidly changing, complex and very competitive environments, where more flexibility, creativity and responsibility are required from individual participants (Grant, 2008). It means, from managerial perspective that more information asymmetries are created, less control is possible and efficient, and large parts of tasks should be left at the discretion of employees. All this makes measurement of the effort exerted by employees more difficult.

Indeed, researchers agree that nowadays a great deal of firm-employee relationships is characterized by the incomplete contracting condition (Gintis, 2005). And exactly under conditions of incompletely specified obligations and only weak or absent performance incentives the effort provided by the employee becomes the key point of concern for the firm and the issue of motivation arises (Gintis, 2005; Gächter and Fehr, 2002).

It is especially true for the jobs in service sector that became a predominant industry in developed countries (Grant, 2009). The contractibility in this sector is relatively low, since the output is highly customized and organizational success crucially depends on the ability of its employees to literally "put themselves into customer's shoes" and come up with personalized solutions. Relationships with the customers become the cornerstone of good performance.

According to Calabrese (2012), productivity represents a more complex contract in service industries than in manufacturing, as there service productivity depends jointly on technical efficiency (internal efficiency) and on perceived service quality (external efficiency). At the same time such sectors need and predominantly employ a new generation of workforce: often better-educated and searching not only for daily bread, but also for daily meaning (Ariely, 2009; Grant, 2009, Terkel, 1974).

Classic motivation theorists and humanistic psychologists clearly supported the notion that individuals have an inherent need for a work life that they believe is meaningful. Maslow (1971) wrote that individuals who do not perceive the workplace as meaningful and purposeful will not work up to their professional capacity. Since meaning is the "primary force motivation in man" (Frankl, 2006, p. 99), meaningful work can be a significant contributor to meeting one's purpose in life (Chalofsky, 2003; Chalofsky and Krishna, 2009). As work constitutes one of the major parts of our lives it is not surprising to suggest that people also seek meaning on and through their jobs.

Nowadays employees are increasingly concerned with benefiting other people and society as a whole (Grant, 2008). According to the literature, many individuals exert high effort doing their jobs not (only) for monetary reasons, but rather because they care about the outcomes of their actions (Prendergast, 2008, Francois and Vlassopoulos, 2008). Such employees can perceive old methods of motivation as too narrow and the organizations sticking exclusively to traditional mechanisms as failing to see the employees as valuable stakeholders with multifaceted needs which should be met.

So the changes in environment for most industries and growth of particular ones (like service sector) along with the changing requirements from customers and employees themselves call for the introduction of new motivational mechanisms able to help organizations overcome the new challenges. However, not only the empirical observation leads us to such conclusion, but also contemporary advances in theoretical field call for the necessity to alter our view of major motives for individual behavior and rethink the traditional mechanisms of job motivation.

2.1.2. Economic research goes beyond selfishness

Contemporary economic science has been enriched by the results of the research in the fields of psychology, sociology, biology and importantly experimental economics. In particular lately the research in the field of economics has been characterized by the introduction of new elements into individual utility functions, namely, social or other-regarding preferences[7] (for standard models see for example Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000). The basic idea here is that people care not only about their personal well-being, but also about the well-being of others. Hence, social preferences become the reasons for prosocial behavior.

Interestingly, researchers believe that social preferences not always come into play. In fact under the conditions of complete contracting, where the rights and obligations of each party under all possible future states of affairs are specified, the policy of carrots and sticks does usually achieve its goals. It is the contractual incompleteness which opens up the room for other motives which can govern behavior (Gächter and Fehr, 2002). As we have argued above incomplete contracting is a common feature of modern organizations, which makes the potential influence of social preferences very possible.

If people are concerned not only with their own material wellbeing, but also care about the wellbeing of others, then this concern can and should be taken into account in designing motivational mechanisms. Organizations should look for the new understanding of employee motivation not only due to the external pressure (competition, customer satisfaction, costs minimization, change and complexity) or internal demand of the workforce seeking growth and self-actualization, but also due to the new theoretical foundation available nowadays, including for instance the theories of other-regarding preferences.

This new understanding that people increase their own utility by increasing the utility of others can be introduced into motivational scheme in many ways. One approach would be to assume that employees can have other-regarding preferences towards their customers and get a boost in their own utility by increasing the utility of the customers. If the satisfaction of customers' needs is a goal of the organization, then creating motivational tools which help the employee to "enjoy" benefiting customer can be very effective.

---

[7] In this manuscript both of these terms are further used interchangeably.

While the concept of social preferences gave a new impetus to the behavioral economic research, there are several issues, which put a limitation on the applicability of this concept in organizational practice. The one of particular interest to us is the context-free approach to the analysis of social preferences adopted in economics. For example Leist (2005) sees the classical view of social preferences as reflecting the "anthropology thesis", meaning that economists[8] try to explain any type of acting in a way that is valid independently of context, which is very problematic.

In our view such understanding of social preferences can prevent researchers and practitioners alike from searching and using the specific contextual tools which could have been used for promoting prosocial behavior. That is why we intend to go beyond the (already well-established) idea of fixed (trait-like) social preferences. We look for the sources of prosocial motivation that can have both other- and self-regarding properties. In particular, the phenomenon of empathy provides us with one possible explanation for prosocial motivation. The self- or other-related image concerns rooted in the existence of internalized or externally imposed social norms can be another reason. Both of these sources of prosocial motivation are highly context-dependent and that is what makes them so interesting for our research.

2.2. A closer look at prosocial behavior

2.2.1. Altruism as a source of prosocial behavior

Every search for the sources of prosocial behavior inevitably leads through the discussion of the concept of altruism. Altruism has been defined quite differently across different disciplines. Many of these definitions originating from economics, anthropology, sociology, psychology and biology are often incompatible. Firstly, altruism can be conceptualized as "helping behavior". That is what is done by developmental psychologists and evolutional biologists. However, even their understandings are not compatible, since the former understand altruism as behavior increasing another's welfare, while the latter see it as behavior reducing one's reproductive fitness (Batson, 1987, 2008). Also, according to Povey (2014, p.5), definitions of altruism which stem from biology are not equate for social sciences in general, as "connection between behaviors which economically benefit others and fitness in biological evolutionary sense is no longer present in contemporary societies".

Economists have long recognized altruism as an important force in economic activity. It was the first and, until fairly recently, only social preference integrated into economics on a wide scale (Konow, 2009). In economic theory, altruism is usually expressed as one person's preference for another person's (or other persons') material or psychic benefit, sometimes called pure altruism (Konow, 2009).

---

[8] In particular Fehr and his colleagues (e.g. Fehr and Fischbacher, 2002, 2005).

Still, altruism is a behavioral phenomenon in economics, which means that the intentions behind apparently altruistic acts are left unquestioned. What matters is the outcome of the actions. One of the reasons for that is that altruism was always seen by economists as a trait: some people just happen to be altruistic. Following this logic, (pure) altruism is often considered as self-sufficient explanation for itself: a fraction of people has a built-in capacity for unconditional other-regarding preferences and get value from giving per se (see e.g. Gintis, 2005, Ariely, 2005).

At the same time economists accept the idea that altruistic behavior in real life can have different motives. Since it is really difficult to distinguish between genuinely altruistic behavior and behavior which appears to be altruistic, Fehr and colleagues (Fehr and Fischbacher, 2002, 2005) explain altruism in behavioral and not motivational terms. Peacock et al (2005) criticize such approach, arguing that, by definition, a really altruistic behavior can be motivated only by altruistic motives while the behavior that appears to be altruistic, almost always has other than altruistic motives. However, one of the biggest problems with unconditional altruism is that people are very rarely indiscriminately altruistic.

2.2.2. From pure altruism to social preferences

The classical altruism was unable to explain multiple phenomenon of human behavior observed in both experimental and real settings. For example, in the ultimatum game people systematically punished those proposers who made them an "unfair" offer, which does not comply with the unconditional other-maximizing preferences of altruists.

In order to overcome the challenges posed by the concept of pure altruism, economists came forth with a solution in form of theories of social preferences. The use of such "social utility functions" marks a convergence between economics and social psychology (Gallucci and Perugini, 2000) and results in the proliferation of the new field of experimental economics. The essence of the new approach to the modelling of human preferences consists in acknowledgement of the fact that the decision maker's utility is a function of both her personal material outcome and the outcome of the other person, with whom the own outcome is compared. This means that under certain circumstances people can prefer less money to more money and more effort to less effort.

According to these theories social preferences is a type of preferences characterized by the fact that "the person not only cares about the material resources allocated to her, but also cares about material resources allocated to relevant reference agents" (Fehr and Fischbacher, 2002). Altruism became just one of the three "quantitatively most important types of social preferences", besides inequity aversion and reciprocity (Fehr and Fischbacher, 2002)

While altruism is still defined as unconditional kindness in such theories, the equity (or fairness) relates to the self-centred preference for equitable outcomes conditioned on the benchmark for such outcome (Fehr and Schmidt, 1999).

According to this new economic perspective, not all the people are concerned about social utility: there is a fraction of people who under any circumstances behave as a classical economic man. For example, according to Engel (2011), the subjects in the meta analysis of 616 dictator games donate on average 28 percent of their endowment. Johnson and Mislin (2011) in the analysis of 143 trust games find that trustors give on average 49 percent of their endowment and trustees return 37 percent of what they received. But there is a considerable heterogeneity of behavior, which is usually explained by the differences in preferences between subjects.

However, admitting the heterogeneity of human preferences, scientists conclude that in addition to purely selfish people, there are others who dislike inequality and are ready to lessen their material payoff in order to restore the equality (Fehr and Schmidt, 1999).

Several models of inequity aversion (e.g. Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000) were more successful than others in explaining such experimental results, showing that people can be altruistic towards others if the other person's payoff is below an equitable benchmark, but the same people can be envious and willing to reduce the payoff of the other person if this payoff exceeds such equitable benchmark (Fehr and Fischbacher, 2002).

2.2.3. Challenging context-free view of social preferences

Experimental economics has recently become one of the most common areas of empirical research in economics, which provides evidence of non-selfish behavior (Povey, 2014). The results of these experiments give valuable insights into the human behavior.

It is important that new models of social preferences explain deviations from selfish behavior not in terms of mistakes or cognitive limitations of humans, but in terms of stable other-regarding preferences (Faillo and Sacconi, 2007).[9] These models often divide people into different types, as for example Andreoni and Miller (2002), who distinguish between selfish, utilitarian and Rawlsian players. The first are concerned only with own material payoffs, the second maximize the total surplus and the third maximize the payoff of the worst-off subjects (Faillo and Sacconi, 2007).[10]

While such view is unquestionably a step forward, it is still follows a behavioral (not motivational) tradition and considers only outcomes of the actions. It narrows the circle of those who can rely of social motives of others, putting fairness as the guiding principle for selection. However, why of

---

[9] However people in these models are still seen as utility-maximizers.
[10] The latter corresponds to the model of "quasi-maximin" preferences by Charness and Rabin (2002)

behavior is still unclear, and behavior itself is set into the context-free frame. This is seen by many researchers as problematic: according to Peacock et al (2005) the definition of altruism should necessarily "capture the intentional dimension of altruistic action" and Leist (2005) believes that altruism cannot be discussed in a context free frame.

If we accept the view of altruism supporting "hardwired egalitarianism" (in the way inequality aversion models suggest), it becomes very difficult to explain, why, for example, charitable giving constitutes very small part of income and inequality in consumption is getting wider (Andreoni and Rao, 2011). People happily coexist with the situation of inequality, and take actions with a cost to themselves only if they are asked to make charitable donation or if they are observed by the audience. Researchers believe that while people have a capacity to behave altruistically, they do not use this capacity most of the time. People feel sympathy in the face of inequality, but their response is mediated by social cues and incentives. As a result selfish behavior predominates. Andreoni and Rao (2011) believe that there are social cues directing altruistic behavior, which are aimed at making this behavior less indiscriminant.

List (2007) shows it using as an example of giving in the dictator game. He provides an analysis of the results for more than 200 dictators and comes to the conclusion that behavior depends on the action set: in particular, the inclusion of "take option" into the choice set immediately makes giving less common. From this List (2007) comes to the conclusion that "behavior is crucially linked to not only the preferences of people, but also the properties of situation". So in a standard dictator game expectation of both givers and receivers coincide, since giving nothing from money one just kindly endowed, is considered very selfish. Such design created the pressure of particular sharing norms. However, if the frame changes and the giver can be non-selfish just by not taking anything from the receiver, the dictator gets "moral authority to give nothing". In fact participants try to use any available contextual clues related to the game to figure out which set of norms can be applied in this particular situation. The results of the List's (2007) study confirm that if taking is an option many fewer dictators are willing to share money.[11]

2.2.4. Complex picture of pro-social behavior

The discussion of altruism in other fields has been based on the *motivational definition of altruism* and, thus, explored the reasons for altruistic behavior. Batson (2008), for example, is defining altruism as a motivational state with the ultimate goal of increasing another's welfare. By "ultimate" here we can understand the goal which is an end in itself. Ultimate goal can be compared, according to Batson (2008), to the so called instrumental goal, which is the one serving the attainment of the ultimate goal. In other words we can talk about (pure) altruism only in case when the wellbeing of another person is the ultimate motivational force for the person's behavior. The reason for such altruism lies in prosocial

---

[11] Although in any way agents do not ubiquitously choose the most selfish outcome.

emotions, and in particular relates to so called "empathic concern", which allows people not only to feel as the other person feels, but also to be willing to contribute to the welfare of the other person in need.[12]

However, other's wellbeing is obviously not the only, and in many cases not the ultimate goal, of apparently altruistic behavior observed in the real life. For example, in the philosophical literature, where altruism is a common concept, there is an opinion that in its pure form altruism can be extended only to the relation within the family or towards friends. In other circumstances people also can behave altruistically, but the reason is "a matter of morality". Altruistic behavior is seen as a part of "good life" (Peacock et al, 2005). So the reason for altruistic behavior would be in this case the internalization of social norms. People behave in accordance with those norms and experience either internally or even externally administered rewards for the compliance and punishment for incompliance with such norms. For Batson (2008), for example, such behavior is nothing more but a subtle form of egoism.

It is not always good to paint everything in black and white, as the truth is usually in-between. We make an attempt to abandon the classical understanding of altruism, which is too "white" and in many cases its standard is unattainable, and move to the understanding of other-benefiting behavior which encompasses different motives (also comparatively selfish ones), each of them being also context-dependent. In this case it really makes sense to analyse those mechanisms which make other-regarding behavior flourish in some circumstances and disappear in others.

Many scientists nowadays, and economists more and more join such position, see pure and impure altruistic motives (i.e. with other's wellbeing both as ultimate and instrumental goal) along with purely selfish motives as pieces of the puzzle which need to be put together to gain more adequate picture of human motivation.

So, Gintis (2005) discusses "prosocial emotions" and internalization of norms along with kin altruism (discussed, for example, in Hamilton, 1963) as the reasons for prosocial behavior. He relates the first two to "characteristically human mental structures". He believes that on the one hand most humans have a predisposition to such prosocial emotions as empathy, sympathy, shame, pride, and spite and exhibit them under certain conditions; on the other hand the society (in form of individuals, like parents, and institutions, like schools) shapes the norms of next generations "in direction they deem desirable".

Ariely et al (2009) have a similar vision and propose three major motives of prosocial behavior: intrinsic, extrinsic and image motivation. Intrinsic motivation includes pure altruism or social preferences (as usually modelled in e.g. Fehr and Schmidt, 1999). Extrinsic motivation for giving

---

[12] There is a heated debate about the role of prosocial emotions and cognitive mechanisms in invoking this type of altruistic behavior and we will touch upon them in the next part of this paper.

comes from material rewards (e.g. in case of charitable giving this can be tax breaks). Image motivation (signalling motivation) is related to the desire for social approval, for being liked by others and oneself (captures rule of opinion in utility and has a lot to do with social norms).

Ben-Ner (2013) classifies preferences into self-regarding, other-regarding and process-regarding with self-regarding preferences being related to the personal self-interest; other-regarding - to altruistic actions; and process-regarding (or social / prosocial) - to trust and trustworthiness, fairness and equity, etc.

According to Batson (2008), the roots of prosocial motivation can lie in pursuing a combination of altruistic, egoistic, collectivistic or "principlistic" goals. While first two types can be clearly understood from the previous discussion, collectivistic goals can be directed to a benefit of particular group and principlism is described as the desire to uphold a moral principle. Batson (2008) argues that both collectivistic and principlistic motivations can be in fact "a subtle form of egoism", because both can serve as the instrumental means for reaching the rewards in forms of higher self- or social-image.

Leist (2005) on the contrary, argues that altruism should be understood as the behavior where "interest in oneself and in others is joined". By that he means that, for example, interest in social recognition can be seen as an integral part of altruistic acting.

In our view such complex view of prosocial motivation leads to the understanding that a wide range of individuals can actually be motivated to behave prosocially. At the same time, there are obviously limitations and boundaries to the contexts where one can observe prosocial behavior. As it has been mentioned above, the fully specified contracts, for example, do not leave any room for pro-social concerns, and in anonymous, competitive settings people often behave selfishly. Also according to Grant (2007) prosocial motivation can be essentially understood as the relational phenomena and should be largely observed in the settings defined in terms of relationships.

2.2.5. Why to search for the motives behind altruistic behavior

The very common question which is attributed to the reasons of "helping behavior" (apparently altruistic behavior which is socially beneficial in terms of Konow (2009)) is why to be concerned, if the positive effect is reached. The main reason is the possibility to intervene and create conditions where such behavior will be more wide-spread. Batson (2008) believes that our behavior is extremely variable and depends on the strength of motives for and against as well as other options available at the moment. Importantly if behavior promotes an ultimate goal (as opposed to instrumental goal) it is more likely to occur.

We follow this view and continue our discussion with the specific goal in mind: as scholars of management we look for particular mechanisms, which can promote socially beneficial other

regarding behavior of individuals, exhibited as part of their jobs. Thus, we are particularly interested in the contextual factors motivating prosocial behavior, the latter being driven either by purely altruistic or impurely altruistic motives.

We look further at purely altruistic motivation and its roots. We start our analysis from the discussion of studies in the field of neuroeconomics, support their results with empathy-altruism hypotheses introduced by Batson (2008) and colleagues, and finally address the research in charitable giving, related to the importance of identification and information disclosure for prosocial behavior.

2.3. Empathy-based prosocial behavior

2.3.1. About empathy and empathic concern

Empathy has been studied for hundreds of years, and still there is no one single concept which could explain the nature of this phenomenon (Preston and de Waal, 2002; Vignemont and Singer, 2006). It has been seen as the source of altruistic behavior by Thomas Aquinas, David Hume, Adam Smith, Charles Darwin, Herbert Spencer and many contemporary psychologists (Batson, 2008).

Major contributions to the concept come from the field of philosophy, psychology and currently more and more neuroscience. Not surprisingly the definitions of empathy are as different as the people working on the topic. However, two approaches are very common in the research: one based on the understanding of empathy as affective phenomena and other as cognitive (Preston and de Waal, 2002). The emotional part is usually associated with emphasizing and cognitive – with so called mentalizing – understanding mental states of others (Fehr and Singer, 2005)

According to Vignemont and Singer (2006) these two should be joined in order to produce an empathic feeling. As a result researchers narrow down the concept of empathy to the states, which are affective in their nature, but isomorphic to another's affective state; which are elicited by observation or imagination of another's state and which are characterized by the fact that the person empathizing knows the source of her affective state.

Empathy in such definition is a necessary but not sufficient ingredient for empathy-driven prosocial behavior. Some researchers (for example, Batson, 2008) believe that in order for helping behavior to occur the perspective taking should be integrated into this process. This perspective taking is an essentially cognitive process potentially leading to helping behavior (when coupled with affective response).

On one hand the perception–action model of empathy (PAM by Preston and de Waal, 2002) states that the observation or imagination of another person in a particular emotional state automatically activates a representation of that state in the observer. This is also supported by the research conducted by

Singer et al. (2004), who found that the empathic response is automatic and does not require any form of engagement of judgment about others feelings.

On the other hand, it does not mean that this activation will necessarily produce the response resulting in the relief of the other's condition. While it can be quite natural to think that empathy and helping behavior are inseparable, the relation between them is much more complex.

A very similar term used to describe the same phenomena is sympathy, but sympathy includes "feeling sorry for" the object. It is less about the physical state of the object, but more about the situation this person is experiencing (Preston and de Waal, 2002).

One of the most influential psychologists conducting empathy-related research is Batson. His papers based on the results of multiple experiments (e.g. Batson et al 1978; Batson, 1983, 1984, 2008) have a profound impact on the understanding of the link between empathy and altruistic behavior. The researcher distinguishes between empathic concern (feeling for others) and empathy (feeling as another person feels). Batson (2008) sees the "empathic concern" – other-oriented emotional response elicited by and congruent with the perceived welfare of someone in need – as the most likely source of altruistic motivation. He includes such feelings as sympathy, compassion, tenderness, and the like into empathic concern.

Batson (2008) admits that even if empathic concern is seen as the reason for helping behavior, we cannot always tell, whether this behavior has really altruistic or selfish routes. For example this behavior can be triggered by desire to reduce aversive arousal (when someone sees the other person in need), or by effort to escape punishment (including self-punishment), or by hope to get the reward. However, Batson believes that while all those motives can be relevant, there are certainly situations where they are not the primarily reasons for behavior, but its unintended consequences.

2.3.2. Moderators of empathy

One of the possible reasons, why empathy not always transforms into empathic concern (and leads to helping behavior), is that it could result in the increased emotional load on the person. Vignemont and Singer (2006) show that people try to narrow their circle of emphatic response. If empathic concern means the desire to relieve the condition of the other with whom one empathizes, then it means some costly actions on the part of the empathizer. Unlimited concern for others is the other term for unlimited (unconditional) altruism, which Anderoni and Rao (2011) as well as Vignemont and Singer (2006) consider not beneficial for human fitness. It is therefore quite natural to expect that people attempt to regulate the scope and strength of empathic feelings towards others.

An interesting aspect of contemporary research on empathy is the attempt to discover the moderators in the link between empathy and empathy-driven behavior. On the one hand there are factors, which

can influence the perception of the other's situation. For example Preston and de Waal (2002) stress in their perception-action model (PAM) the importance of interdependence or interrelationship, which lead to the higher chances that the subject attends to the state of the object. They also emphasise the importance of past experience, learning and cue salience. Interestingly, among the most important factors influencing the perception in the PAM are **familiarity** and **similarity** with the object.

Vignemont and Singer (2006) also propose a contextual approach to empathy. In particular, they argue that empathy can be modulated by appraisal of the situation, and this modulation is not necessarily explicit, but can be fast and implicit. In their research Vignemont and Singer (2006) distinguish between two types of modulation of empathy: voluntarily (some people, like medical practitioners or Buddhist monks can intentionally control their emotional responses) and implicit appraisal processes, which might strongly influence the magnitude of empathic responses. The researchers distinguish between four major categories of modulatory factors: characteristics of emotions, nature of the empathizer herself, situative context and **relationship between emphasizer and target**.

This stress on relationship within the dyad and familiarity between them is not very surprising if we remember the well-known biological explanation of altruism, which goes back to the works of Hamilton (1964) who posited that altruism depends on relatedness of the genes between giver and receiver. Hamilton's idea of kin altruism assumes that the closer people are the higher is the probability of altruistic acts of one of them on behalf of the other. While in relations with strangers kin altruism does not work, we probably unconsciously still see the people familiar to us as closer to us and hence, deserving more chances of altruistic behavior.

Thus, in modern human societies "kin selection is likely overgeneralized to the point that simply feeling closer to someone evokes a similar emotional and behavioral response – regardless of any genetic relationship" (Small, 2011, p.151).

The support for importance of familiarity for the other-regarding behavior comes also from the field of behavioral studies, which explain the effect by the cognitive limitations of humans. When people communicate with strangers they often use heuristics to reduce the cognitive load and achieve more effective decision-making. One of such heuristics is that we tend to like those people with whom we already have some relationships. In this case people are readier to comply with requests. So when requests come for example unexpected and face-to-face, people search clues for taking the decision. As researchers found out, giving people the opportunity to see each other – even in the absence of verbal communication for a short time substantially increases compliance with this person's request (Cialdini and Goldstein, 2004).

What makes people familiar with each other, or when do they have a relationship? Anonymous unidentifiable strangers can rarely invoke empathic feelings. Communication is the cornerstone of relationships. People need information about others to let themselves experience empathic concern.

Starting on the one end of the continuum with silent identification and moving to the contexts with one-sided communication to mutual face-to face-communication we can observe how empathic feelings evolve, develop and lead to other-benefiting behavior.

2.3.3. Context for generosity: Evidence from the field of charitable giving

One of the fields extensively involved into the research on empathy and contextual factors influencing helping behavior is charitable giving. Researchers in the field strongly believe that although such motives as social norms, perceptions and religious beliefs can also play a role in giving and helping behavior, it is the basic wish to help people, particularly when those people are in need of external help, which is the core motivation for helping (Kogut and Ritov, 2011).

However, the researchers admit that there are huge disproportions in where and when people help. The major task for the scientists in this field is to understand why people are ready to open their hearts and pockets in one situation and are reluctant to do this in another, and derive applicable solutions for the practitioners.

One of the leading themes in this research is closely related to the ways of decreasing social distance between the giver and receiver. For example Small (2011) believes that the most important determinant of helping behavior is social distance. Factors that reduce social distance promote sympathy and prosocial behavior. Among those factors are perspective-taking, identifiability and similarity. The basic proposition is that when contextual factors reduce the social distance, the sympathy arises, and in its turn promotes helping behavior.

Perspective taking has two sides: either imagining how the other feels or imagining how one would feel in the same situation. In both cases there is a sympathy as a result, but in the latter case there is also personal distress involved, which consequently moves the focus to the self. So if helping is not seen as a way to provide a relief, the motivation for it can be then very low.

Dickert et al (2011, p.165) on the contrary proposed a model which distinguishes between two affective reactions to the person in need: mood-management and empathy. Results of their experiments supported the proposition that "people's own mood management is an important motivation to initiate helping, presumably because the negative feeling that arises when witnessing the suffering of another provides a strong impetus to leave the negative arousal state behind by making donation." So according to Dickert et al (2011) those who experience strong negative arousal are more likely to give. However, the amount of giving is linked to the degree to which the donor empathizes with the recipient.

Gehlbach et al (2012) in their attempt to understand what initially motivates individuals to take the perspective of others also employ the concept of perspective taking. In particular they see the social

perspective taking (SPT) as individual's motivation and ability to discern the thoughts and feelings of others. Although SPT is commonly viewed as the ability to understand people accurately, it also includes the motivation to employ this ability. In other words one should be motivated to try to understand the other person and then she must actually engage in the process which allows to accurately evaluate the other's mental state. In general SPT overlaps with empathic accuracy, but SPT stresses the combination of ability and motivation.

Researchers use the framework of school education and as a result of a study, which included a survey, a performance task, and in-depth interviews with teachers and students come to the conclusion that using specific tools for motivating perspective taking is a "more promising pathway for bolstering people's engagement than trying to change people's general disposition to engage in perspective taking" (Gehlbach et al, 2012, p.12).

Identifiability of the recipient is also an issue addressed in many studies (Kogut and Ritov, 2011, Small, 2011). The major argument goes back to the research of Schelling (1968), who distinguished between individual life and statistical life, arguing that people are willing to extend more help to identifiable victims. Although this effect represents a departure from economic rationality, Schelling's (1968) conclusions have been supported by the later experimental research, which showed that people are more generous to those whom they can identify.

Importantly, when talking about identifiable victim effect, Cryder and Loewenstein (2011, p.237) stress that this "general phenomenon is not limited to the cases involving donations" and cite the research by Redelmeier and Tversky (1990), which showed that physicians treat individual (identifiable) patients more attentively and generously (e.g. recommend in-person visits versus. phone consultations), than (average, statistical) patients in general.

In their quest to understand the reasons for such behavior, researchers came up with several explanations. Firstly, there is so called singularity of the recipient. In particular Kogut and Ritov (2011) argue that the processing of information related to the singular individual is different from that of the group. While single individual is seen as "psychologically coherent unit", group is not. When facing the individual need people process information in a faster and more extensive way. And the more information people have about the individual (name, pictures, etc.), the more they are willing to help a single individual rather than a group (even if they possess the same amount of information about each member of the group).

Kogut and Ritov (2011) conducted several field experiments with charitable donations to a single recipient or a group of recipients. Providing info about individual while asking to help her (sick child) resulted in significantly more donations than to unidentified recipient or a group of recipients, while providing the info about a group of eight recipients did not result in more contributions than to the same group but unidentified recipients. This results have supported the proposition that identification

activates an emotional response, where single person evokes stronger feelings and hence willingness to contribute.

As Small (2011, p.150) points out "in contrast to cold cognitions, which are relatively unchanging and unobtrusive, emotions are passing and signal to an organism to stop current activities, pay closer attention, and respond to the emotion".

Interestingly when people are openly confronted with the choice whether to contribute to a single individual or to the group, they contribute more to the group. This fact also supports the idea that "single victim effect" has emotional roots. However, in case of such "debiasing" the level of willingness to contribute to the identifiable recipients goes to the level of unidentified ones not the other way around.

While identification of a single recipient leads to higher willingness to help, there is a number of factors which moderate the relationship. For example belonging to the giver's group[13] is one of such factors. Social categorization as one of the in-group can promote the feelings of greater closeness (Kogut and Ritov, 2011). This factor is very close the similarity, which is according to Small (2011) influences the helping behavior through by reducing social distance between counterparts. Small (2011) defines similarity as a feeling of closeness to people with whom one has something in common.

The research of Kogut and Ritov (2011) - study of donations after earthquake in the Indian Ocean in December 2004 - showed that increased willingness to help identified victims was relevant only for those situations where they belonged to the same nationality as donors. In this case identification of a single victim worked well: 48% contributed to the identifiable victim and 23% to identifiable group. But in other-nation condition the situation was reversed: 10% contributed to the identifiable victim and 22% to identifiable group. The same relationships were true for the amount contributed.

While nationality can be crucial for considering a person in-group versus. out-group, other factors like religion, ideology can also play a role. Art tastes for example can be a factor for such classification. Experiments showed that identifiable[14] in-group recipient in the dictator game received on average 5.4 (out of 11) shekels, while unidentified in group got only 2.6. As for out-groups there were no differences due to the identification (3.3 versus. 3.4 shekels) (Kogut and Ritov, 2011).

However, when such factors as political beliefs are highlighted in the information about the recipient, the situation can change and, as the experiment by Kogut and Ritov (2011) showed, result in a higher contribution to the unidentified in-group recipients than to the identified and vice versa for out-groups. The authors claim that nationality is not a dominant factor in information, but the one helping to

---

13 Close to the notion of similarity in Vignemont and Singer (2006)

14 It involves weak identification. It has been applied before in Small and Loewenstein (2003) in Dictator game experiment. In both cases participants know that they will never learn the real identity of the recipient. In Small and Loewenstein (2003) identified recipients received on average 60% more donations than unidentified.

strengthen identifiability and singularity. Ideological label, on the contrary, is a dominant factor, which strengthens the emotional arousal towards out-groups and which can be eased by the affiliation provided by identifiability and singularity.

Also according to Small (2011) portraying an identifiable victim expressing sadness is more effective than portrayals expressing happiness or neutral emotion. Small believes that it happens due to the fact that emotions are contagious, and shared emotion encourages the higher empathic response.

At the same time sympathy can be promoted by right framing. For example all the above factors are based on the perception of the givers (how similar they perceive the other, how close, etc.). So if one draws attention to some similarity, even superficial and arbitrary (art tastes, clothes, etc.), it can result in higher chance of helping.

Another very popular idea in this stream of research is the importance of so called tangibility for motivations prosocial behavior. Tangibility, according to Cryder and Loewenstein (2011, pp.237-238) "increases the perception that one's involvement will make a difference" and "intensifies emotional reactions".[15]

Cryder and Loewenstein (2011) believe that tangibility is one of the explanations for identifiable victim effect. They see identifiability as a special case of tangibility and argue that identifiable recipients are inherently more tangible than statistical and abstract ones.

However, there is the connection between tangibility and generosity beyond the identifiability. In a set of experiments Cryder and Loewenstein (2011) tested the positive relationship between tangibility and generosity. They used choices between donations to different charities, letting the donors choose the charity either before or after the choice of the amount to donate. Those donors who had first chosen whom to donate to and then how much, donated more. Those who donated to a particular charity reported greater feeling of impact. Also providing the more tangible information about the recipient (charity) in form of concrete project supported by donations, resulted in almost doubling the amount of donations to the same charity with abstract and general information attached. The results were confirmed in another real-world experiment, where participants donated to the charity with tangible task almost three times more to the "intangible charity".

Also researchers showed that "the increasing feeling of impact not only leads to greater giving, but also leads to greater emotional satisfaction *from* giving" (italic original) (Cryder and Loewenstein, 2011, p.243).[16]

---

[15] Tangible information, broadly, is information that is specific and concrete as opposed to general and abstract (Cryder and Loewenstein, 2011, p. 241).

[16] That's the reason why people are more willing to donate in the end of the campaign than in the beginning. Goal proximity increases the feeling of impact. Since efforts at the end of the process are perceived as more concrete and influential as the ones in the beginning, that why increasing feeling of impact leads to higher satisfaction from giving. In anticipation of that people have done more.

Thus, the tangibility (concreteness) of contribution positively influences feeling of impact and also leads to stronger affective reaction (sympathy) with the recipient. Also the very feeling of impact from giving can directly produce affective reaction in form of warm-glow (will be addressed in further chapters). Both feeling of impact and affect, according to the model of Cryder and Loewenstein (2011), increase generosity.

Authors believe that in real world the attempts to increase the concreteness of goals and encourage affective reaction can increase contributions. This is very much in line with the conclusions of Dickert et al (2011) who admit that it is very important that donor believes that her contribution can make a difference. Then she can contribute at least some amount. Further Dickert et al (2011) states that in order to increase the amount people should be primed "to feel rather than to think", since when givers have a chance to take the perspective of the recipient (get into her shoes), they could empathize more.

Cryder and Loewenstein (2011), unlike Dickert et al (2011), argue that taking the decision on how much to give based on emotional reactions is not efficient, as it leads to inefficient distribution of resources. So in the prescriptive tone they advise that " relying on our sympathy for decision about *whether* to give and relying on our reason for decisions about *how* to give may yield the best policy of all" (italic original) (p.248).

2.4. Image-based prosocial behavior

2.4.1.   From "warm glow" to the importance of norms

While the research in other-regarding behavior rooted in empathic feelings for others argues for unselfish motives for such behavior, there are apparently other less unselfish motives driving seemingly prosocial behavior.

One of the most prominent researchers working on the topic of non-altruistic motives for apparently altruistic behavior is Andreoni (1990, 1998), whose main area of work is related to the analysis of charitable giving - the field where one can expect the prevalence of pure altruistic motives.

One of the most important observations made by Andreoni (1990) is that the large part of charitable donations is induced not by the purely altruistic motives, but by the experiencing of so called "warm glow". Warm glow in this context is defined as the increase in utility resulting from the very act of giving additionally to the gain in utility from increasing the total supply of privately provided public goods. The combination of pure altruism and warm glow is referred to as "impure altruism" (Andreoni, 1990, 1998).

Why do people experience warm glow? One of the explanations is that giving behavior is often associated with following moral norms. Behaving in accordance with some "higher principle" (what Batson (2008) calls "principilism") can bring a utility of its own, which is the warm glow.

Benabou and Tirole (2006) conducted a very interesting experiment aiming at separating the effects of pure altruism and warm glow. In their experiment all subjects were getting a fixed compensation and also had an opportunity to contribute by their effort to the charity of their choice (excluding the subjects in the baseline treatment). For the first group independent of the effort the difference between own contribution and maximum donation was covered by the experimenters (thus fully crowding out the purely altruistic motivation) and for the second – contribution to the charity fully depended on the effort of the participant. Thus, any additional effort which can be observed in the second case can be attributed to the pure altruism (which they call output-oriented altruism), while the effort in the first treatment can be seen as the result of warm glow (which they refer to as action-oriented altruism).[17]

The results supported the importance of action-oriented altruistic motivation in the workplace (rise in productivity of 15% compared to the baseline treatment), while the output-oriented motivation proved to be insignificant.

Tonin and Vlassopoulos (2009) also conducted a similar experiment to distinguish action- and output-oriented altruism. The experiment involved a real-effort task (introducing data in Excel format). The researchers confirmed the results reported by Benabou and Tirole (2006): action–oriented altruism accounted for significant increase in effort, while output-oriented altruism did not provide any additional impact on the participants' level of effort.

The warm glow effect has its roots in the desire to follow the implicit or explicit social norm even if the act of compliance with the norm has been unobserved by other (we know that social norms include non-monetary rewards and punishment by others). In this case researchers talk about the motivation based on compliance with internalized social norms.

The moderators of such behavior have very individual specifics – even if moral norms are universal adherence to them can vary across individuals due to such personal factors as upbringing, cultural background, etc. Since the rewards and punishments in case of moral and internalized norms are internally administered, it is very difficult to determine the clear link between the norm and observed behavior. This is, however not true in case of explicit social norms, where external social pressure is shown to be one of the main determinants of behavior.

---

[17] Action-oriented and output-oriented altruism are "logical counterparts" of warm glow and pure altruism (Tonin and Vlassopoulos, 2009).

2.4.2. Social norms in economics and sociology

Social norm can be defined as a behavioral regularity that is based on socially shared belief of how one ought to behave, which in its turn triggers the enforcement of prescribed behavior by informal social sanctions (Rauhut and Winter, 2010). It is important that sanctions should not necessarily be costly for the punisher, as even indirect social sanctions (born out of suspicion that someone dislikes her behavior) can be very influential (Rege and Telle, 2004).

Akerlof (2007, p.8) points out that classical utility functions are described too narrowly, since "they depend only on *real* outcomes" (*italic* original). In reality people have views about how they should or should not behave, which can be called norms. Standard economic utility functions miss this important aspect of motivation.

There is a considerable difference in conceptualization of human behavior by sociologists and economists (Rauhut and Winter, 2010). Sociologists make emphasis on social norms and see the individual as homo sociologicus, "who is a pure marionette of normative and role expectations" and is apparently oversocialised. Economists, on the contrary, have their homo oeconomicus, who is egoistic maximizer. Homo oeconomicus can be seen as "undersocialized". Only recently there started research trying to combine both approaches.

Some signs of such merger can be found in the theories of social preferences. Normative concerns (e.g. equity) can be seen as a basis for the theories of social preferences, but these theories have limitations based on the fact that they relate preferences only to outcomes of the actions. The motivation comes from the combination of self-interest and other-regarding preferences in the form of equity concerns (models of Fehr and Schmidt, 1999 or Bolton and Ockenfels, 2000) or quasi-maximin preferences (models of Andreoni and Miller, 2002 or Charness and Rabin, 2002). While these models are without questions very novel as they include the concern for other individuals into utility functions, they still do not talk directly about social norms (Faillo and Sacconi, 2007). The decision of individual to comply with the norm is derived indirectly from her affectivity towards the outcome. Her preferences seem to be independent of the degree of compliance of others with the same norm. The same is basically true for the reciprocity models, which assume that "what matters is the kindness that each player expresses directly towards her opponents, not compliance with impartial norms which could involve any other subject" (Faillo and Sacconi, 2007).

Leist (2005) also criticizes the explanation of results of economic experiments by purely altruistic motives without taking into account the influence of social context. He argues that claiming (as e.g. Fehr and colleagues do) that fair acting is not influenced by social context, is equal to claiming that "moral equality is an anthropological tendency of behavior and not a cultural achievement" (Leist, 2005, p.11).

Ellingsen et al (2012) distinguish between three types of social framing theories: variable social hypothesis (frames affect internalised norms or preferences); image hypothesis (frame affects how others interpret one's behaviour, relates to social esteem) – here even selfish people may want appear prosocial; coordination hypothesis (frame affects expectations which people have about each other's behaviour).

2.4.3. Putting norms into context

There is mounting experimental evidence (e.g. Faillo and Sacconi, 2007, Bicchieri 2006, Bicchieri and Chavez, 2010) for the fact that people can conform to the particular norm in one situation and violate it in the other. In such cases it is important to concentrate on the contextual factors which promote or prevent conformity. According to Bicchieri and Chavez (2010) when people encounter particular situation, they do the categorization of this situation and roles and norms acceptable in this situation. This process depends on several factors, such as past experience, individual's goals or framing (how the situation is presented). In other words, contextual clues are taken into account in the process of categorization of situation which one is involved into. This categorization enacts particular norm which in its turn influences preferences and beliefs.

So while in economics it is usually assumed that people know their preferences and take decisions on the basis of these preferences, sociologists assume that in real life people use social and moral norms to guide the behavior. This means that the utility function depends on the "degree of one's adherence to the perceived norms of others" (Charness and Schram, 2012).

Sociological research addresses several aspects of norms: in particular, conditionality (either the norm holds under all circumstances); intensity (degree to which individuals subscribe to the norm); and consensus (extent to which the norm is shared by the members of the society). In fact in the economic experiments those properties of norms are often not considered. For example conditionality first of all means that norms can depend on the characteristics of the situation or the individual. So equity norm for example is a conditional norm (depending on the relative amount of effort one put into the project), while equality is an unconditional norm which does not depend on distributional concerns like need, status or effort (Rauhut and Winter, 2010).[18]

For example Grimalda and Sacconi (2005) believe that there are two major classes of motives which can explain individual behavior: consequentialist (mainly self-interested) and conformist. For conformist preferences the situation is described as either conforming or not to the given abstract principle or norm or ideal of fairness. At the same time conformist preferences of the players depend

---

[18] Indeed, when talking about the role of norms in behavior, researchers often attribute the explanation to the rules of distributive justice. In particular they claim that such principles as need, equity and efficiency lie in the foundation of giving behavior, observed in the laboratory and in the field.

on her beliefs about the conformist preferences of other players. Simply put, player as much complies with the norm as she believes the others will do.

Individuals often turn to social norms when they need to gain an accurate understanding of social situations and properly respond to them. It is common to distinguish between injunctive norms, which inform about what is approved/disapproved and descriptive norms, which relate to what is typically done by people. But the extent to which these norms are taken into account and followed crucially depends on them being focal and being in alignment with each other (Cialdini and Goldstein, 2004).

Basically people conform to the norm only when these norms are made salient, in other words when they are in focus (via environmental cues, which focus the attention on the norm). In order to follow norms, the normative information should be "highlighted prominently in consciousness" (Cialdini and Goldstein, 2004). [19]

2.4.4. Observability as a factor increasing the salience of norms

People are social creatures and living in society they developed rules of social co-existence. They want social recognition and dread social astorcism (Povey, 2014). Having meaningful social relationships with others is one of the basic human motivations. Also people believe that if they do what others approve of, then these others approve of them too. People follow the shared social norms, as they expect to be approved for complying with those (Cialdini and Goldstein, 2004).

Benabou and Tirole (2006) formalized the reasons for norm-adhering behavior in their model of social signalling. The researchers used the results of their experiments, separating the effects of action- and output-oriented altruism, to support their so called signalling model. In this model they postulate that people's actions reflect a mix consisting of material self-interest, altruistic motivation and social or self-image concern (which they call "social signalling"). They believe that this mix can vary across individuals and situations, and it can be quite difficult for the observer (or even the person herself) to understand the person's true values. Interestingly they stress that changing any of the three motivational components can alter the meaning given to the actions and consequently incentives for performing this action.

The signalling model of Benabou and Tirole (2006) has been tested by Ariely et al (2009), who, following the premises of the signalling hypothesis, suggested further that people act more prosocially in the public sphere than in private settings. They argue that image crucially depends on visibility of actions, which means that changing visibility changes the level of prosocial activity. In order to signal

---

[19] The explanation comes from the "focus theory of social norms" born out of "littering experiments" by Cialdini et al, 1991. Cialdini and Goldstein (2004) admit that the major challenge in this respect is to make the norms salient also in the long run, as normative information becomes less accessible over time.

traits defined as "good", one should base her behavior on common shared social norms. Thus, prosocial behavior is "a way to signal to others that one is good".

Their major hypothesis was therefore that increasing visibility increases the level of prosocial activity. In the experiment "click for charity" they have learned that when the effort of the participant was made public subjects exerted significantly more effort supporting the good cause (Ariely et al, 2009) The researchers concluded this "dependence on visibility" to be a unique characteristic of image motivation.

Similar results were earlier reached by Rege and Telle (2004) whose experimental study explored how social approval affects cooperation. It is a common approach in economic research to assume that people have preferences for social approval, but the choice is not explicitly modelled. The experiment by Rege and Telle (2004) made the social approval more explicit.

The results of experiment clearly supported the hypothesis concerning the contributions under the pressure of indirect social approval. Importantly this happened among strangers. The experimenters just asked participants to contribute openly (everybody could see how much each participant contributed to the common box). Under such treatment the average contribution increased from 34.4% to 68.2%.

Thus, social norms have power because people seek approval from others; hence, the observation is crucial for functioning of social norms. This acceptance and approval by the group is very important for self-identity. Social norms work through such emotional mechanisms as shame, which also means that they are mostly acute when one's actions are externally observed (Charness and Schram, 2012).

2.4.5. Lessons learned for our research

Summing up the results of the research in different field we come to the conclusion that the frame in which the phenomenon of altruistic behavior is discussed should be broadened to allow a wider spectrum of other-regarding behaviors and their motives. While critique of pure altruism and its inability to explain real facts about human behavior has led to introduction of social preferences, the latter also suffer from the same limitations, the major one being its context independence.

Current research in neuroeconomics, psychology and charitable giving provide us with alternative account of information about human propensity of exhibit other-regarding behavior. In particular we learn that altruistic actions result from our ability to experience empathy and empathic concern. While this ability is universal across humans, they have an ability to regulate (both consciously and unconsciously) their emotional response, which many researchers see as a healthy adaptation strategy. One of the essential positions among the adaptation mechanisms is taken by the familiarity with the

subject. In other words in order to go from empathy to action people need to identify with the other and this is hard to achieve with anonymous unidentified beneficiary.

Identification is also shown to be one of the major factors promoting contributions in the context of charitable giving. For our research it means that employees can behave in a prosocial way when they have an option to perform the tasks for identifiable customers.

Another reason for prosocial behavior is the desire (or necessity) to follow shared social norms. Adherence to the norms is not universal and depends both on the characteristics of the norm (conditionality, intensity and consensus) and on the situational context (e.g. saliency of the norm). However, one of the major contextual factors influencing norm adherence is the observability of the actions. In our view this effect of observability can be also seen as an identification effect: if the giver can be identified with her decision, then the option of providing this information to other external parties can trigger the norm-related behavior from the giver.

In the context of employee-customer relationships this link can be emphasized when the employee is aware that customer has information about her and can match the quality of the product (service) she gets with the effort of particular employee. We believe that even in the non-strategic settings where there is no direct reputation building effect such awareness can trigger change in employee's behavior. It is important though, that this is only true for the context where explicit social norms promoting employee's effort are salient.

2.5. Power of context in prosocial behavior

2.5.1. Evidence from experimental research in economics

The importance of contextual factors and in particular identity disclosure and communication has found support in recent experimental research in economics. While traditional economic experiments have a common feature of securing anonymity of subjects and not allowing any forms of interaction (which allows tighter control), such approaches have been criticized for their lack of external validity, as such fully anonymous situations in many cases do not reflect the real-life interactions.

That is why a part of researchers has made an attempt to introduce social interaction into experimental settings, with prominent researchers in the field recently starting to accept the necessity to introduce more realistic settings into the experimental design (e.g. Bohnet and Frey, 1999; Andreoni and Rao, 2011, Andreoni et al, 2012). In particular, researchers in multiple experiments allowed their subjects to conduct non-binding communication at the pre-play stage of the experiment and came to the conclusion that such communication dramatically increases the subsequent cooperation between the subjects.

However, communication is a very powerful tool involving both parties. Bohnet and Frey (1999) argue that communication is not the only moderator mechanism. Silent identification can be enough to change the course of subjects' behavior. They state: "In the dictator game, one-way and two-way identification induces more `other-regarding' behavior. Our experimental results thus indicate that removing anonymity suffices to increase solidarity. The `Sound of Silence' has a power of its own" (Bohnet and Frey, 1999, p.44).

Researchers argue that by identifying another person one can reduce social distance which allows higher level of empathy. "Standing by and neglecting a specific child who has fallen into a well is much more difficult than not rescuing an unspecified statistical life, for example, the anonymous children starving from hunger" (Bohnet and Frey, 1999, p.46).

Bohnet and Frey (1999) in their experiment allowed for one-sided identification, which in their words led to higher solidarity and higher propensity to help. The experiment involved two treatments differing in the amount of information provided about the receiver: one with silent identification only and another with provision of some information. The results showed that the more the dictators new about the receivers the more generous they were. The experimenters thus came to the conclusion that solidarity increases with decreasing social distance and even in case of one-sided identification. It helps to transform "abstract, anonymous stranger into a visible, specified individual. The closer are the recipients to potential benefactors, the more the latter value the recipient's well-being. It could be argued that "the other's well-being is discounted depending on social distance as future benefits are discounted depending on temporal distance" (Bohnet and Frey, 1999).

Charness and Gneezy (2000) in their review of the literature devoted to the influence of communication on prosocial behavior cite many researchers demonstrating that such behavior is activated in non-anonymous settings and increase when the social distance between subjects get reduced. As Bohnet and Frey (1999) posit, fairness considerations are not always active – there are some characteristics of situation which make them so. When the subjects are interacting in anonymous environment, they have only a purely intrinsic motivation to behave fairly; with identification fairness becomes partially activated; but when people also communicate with each other, the fairness norm is strongly active. Identification can provide subjects with information concerning gender and ethnicity and thus affect behavior.

Charness and Gneezy (2000) emphasise that many field interactions are conducted with neither complete anonymity nor complete familiarity. When people are involved in interactive situations, they usually have some clues concerning the characteristics of others. So the researchers test their hypothesis in the dictator game and the ultimatum games. In the control treatments they use the standard set-up of anonymity only with respect to the other participant(s). As a treatment variable, the participants also learn the family name of their counterpart.

The hypothesis tested is whether this additional information, regarded as irrelevant by traditional game theory, affects behavior simply by reducing the social distance between participants.

In the dictator game, they find that providing family names results in more generous allocations. However, in the ultimatum game revealing the name of the recipient had no significant effect on behavior. From that researchers conclude that social preferences may have more influence in situations that are perceived to be less strategic in nature.

At the same time many experiments show that behaviour in the multi-person games respond to the name of the game / strategy labels. Dreber et al (2011) argue that the results of economies games do not reflect the change in preferences, but the change in beliefs, as social context serves as a coordination device entering not preferences but beliefs (about opponent's behaviour).

While other researchers say that minor changes in how the situation is described can change the preferences of participants (and use Prisoner's dilemma for testing the proposition), Dreber et al (2011) conduct a series of dictator games (three large experiments) with different game names (framed as taking game or giving game) or recipient's knowledge about the game and come to the conclusion that there is no difference in behaviour. In dictator game the particular frame, they argue, can affect only preferences and not beliefs (about other's behaviour). Thus, a lack of significant difference in behaviour is taken by researchers as a supporting argument for the proposition that frames change only beliefs.

In an attempt to explain the reasons behind the change in beliefs Dreber et al (2011) turn to the norms governing behaviour. They assume that in the Dictator game there is less ambiguity about the social norm than in Ultimatum game: norm is stronger when one faces powerless opponent, while when the opponent has at least some power proposers feel more competitive and less socially responsible. That is why framing matters less. Also in the ultimatum game there is an opportunity for the responder to show her reaction to the behaviour of the giver at least in the form of rejection. Even in the Dictator game possibility of feedback has sizable impact on allocations. For example, communication (in form of offer rejection or in form of a message) stresses the role of social norm in the dictator's mind and in this case labelling the game can have a role. Interestingly researchers suggest that possible mechanism, which accentuates norms, could lay in the reduction of social distance.

According to Andreoni and Rao (2011), "in the real world giving occurs in the context of social interaction between giver and receiver and the incentives to give are affected by social factors surrounding the interaction".

The experiment conducted by Andreoni and Rao (2011) distinguished between the one-sided communication (either on the side of giver or receiver) and two-sided communication. They found out that when the giver had the exclusive opportunity to speak, only 6% of endowment was allocated (this

is much lower than even in no-communication condition where 15% was donated). If the recipient was the only to speak, at least 24% of endowment was past to her. The highest, however, was giving in the condition where both could speak – 30% on average (even if allocator was speaking first).

Authors address such difference to the "power of ask", which they found in the analysis of the messages content. They discover that allocators were more often talking about fairness when they were asked by receivers (even if they did it simultaneously with receivers not becoming but just anticipating the message). In case when only allocators could speak, messages were typically offering apology. Authors link their results to the research in neuroeconomics and come to the conclusion that "communication from the recipient facilitates altruism through feelings of empathy". They understand empathy as the allocator's readiness to actively consider and identify with the role of recipient. However, they admit that there are also selfish motives which can help explain seemingly altruistic behavior, e.g. guilt aversion, self-image and self-signalling and self-deception. Authors believe that self-image explanation is also well suited to their results, as communication (on the side of recipient) forces the giver to know and consider the receiver's feelings (as opposed to the potential knowing such feelings in other conditions). Hence, it is becoming more difficult for the receiver to use self-deception. At the same time as self-deception and self-signalling are becoming more costly, the empathic concern raises and promotes altruistic behavior.

Authors do not give a clear answer to the question whether altruistic behavior based on the empathic concern is motivated by pure concern about others or by selfish desire to reduce the cost of self-deception. But in their own words communication is one of the mechanisms that "help us through the empathic feelings remove the blinders protecting us from naturally altruistic tendencies".

As we see the results of the above research support the idea that identification and communication on the part of the receiver triggers affective reaction of the giver, promotes empathic concern and results in higher propensity to exhibit altruistic behavior.

2.5.2. Evidence from research in organizational science

In organizational studies agency theory remains the dominant theoretical framework associated with performance management. Its key premise is that principals and agents have different preferences and agents exploit information asymmetry to engage in shirking or pursue their own preferences. Meanwhile agency theory was broadened by incorporating of the possibility that agents may be not entirely self-interested and can respond to the appeals to shared beliefs about the value of the services provided by the organization. However, these developments "stayed within academia rather than shaping reform efforts" (Moynihan et al, 2012).

In general performance measures are essentially important in solving the information asymmetry problem and this led to their wide acceptance in different settings. At the same time agency theory sees the agents an opportunistic and self-interested. According to Moynihan et al (2012) the most extreme failures of performance system were observed in the cases where principals closely followed the prescriptions of the theory and relied on highly-loaded incentive measures, but could not prevent agents from "gaming these measures for perverse ends".

While agency theory has largely ignored other than self-interested sources of motivation, research in the fields of public administration, political science, and organizational theory documents the potential of prosocial values.

While the idea of prosocial on-the-job behavior is not common in classical economics, it found some support in field of public economics, where it has been often studies through the prism of such parameter as "job impact". According to Francois and Vlassopoulus (2008), workers in the field of education, health care, child care, charities, and social work often explain their motivation as a "desire to make difference", i.e. to bring positive change. About half of all paid charity workers state that they are driven by mission and that is what makes them agree to a lower pay: they have a chance to "help people and make a difference".

Mission consists of attributes of the project that make some principals and agents value its success over and above any monetary income they receive in the process (Besley and Ghatak, 2005).

Thus, there is a significant evidence that individuals with greater desire to serve others are more likely to work in public organizations due to the opportunities to provide meaningful service and also because they find such work rewarding. The benefits of such motivation depend on the degree to which employees feel that their work indeed provide them with opportunities to serve others. Such conclusions are consistent with expectancy theory (Vroom, 1964) and goal theory (Wright, 2003) stating that employee outcomes are contingent on satisfying employee expectations and goals. This view however assumes that employee has (stable, inherent) public service motivation.

Besides extrinsic rewards, an important drive is a concern towards the social cause pursued by the organization people work for, or a sense of altruism towards the welfare of a third party that is the recipient of the good or service being produced in their workplace. Such workers are willing to make labor donations, by providing on-the-job effort beyond what is contractually required of them. There is mounting empirical evidence that this type of labor donations are important in organizations engaging in the provision of education, health care, child-care, and social services as well as in charities and NGOs that advance all sorts of social missions (Tonin and Vlassopoulos, 2009, 2012).

For example Moynihan et al (2012) propose an alternative (to agent) theory of performance management that rests on prosocial values. They choose the parameter of performance information

used by public servants as the dependent measure. Public servants who see the social impact of their work, are more likely to use performance information (it is used for improving program performance and for presenting evidence for their performance), which leads to improvement of services and maintain the resources. They support their theory using the results of analysis based on cross-sectional survey of US public and non-profit employees.

Researchers conclude that building a culture that rests on the notion of meaningful significance of organizational goals, relying on the repeated use of appropriate symbols, can promote norms that centre on social impact.

In general recent burgeoning theoretical literature in economics recognizes the important role of workers' pro-social motivations and examines their implications for the design of incentive contracts, the selection of workers, the provision of effort and organizational design. For instance, Besley and Ghatak (2005), Delfgaauw and Dur (2007, 2008), Dixit (2002), Francois (2000), Glazer (2004), Murdock (2002).

In the public economics literature the stress is made on the view that if employees agree with the mission of the organization they can provide a level of effort similar to private organizations, but for significantly lower remuneration. A traditional explanation rests on the assumption that workers care about the mission of their jobs. The prevailing opinion in the literature is that matching the employee with the appropriate mission happens at the selection phase. That is why lower powered monetary incentives and specific modifications to the contracts are offered (e.g. Francois, 2000; Dixit, 2002; Prendergast, 2007).

In general mission is not only about what firm does, but also how (e.g. environmentally friendly), who the principal is (residual claimant, government), etc. Although public sector is usually seen as mission-driven, private firms frequently adopt strong mission, e.g. Body Shop, Mary Kay, Wal-Mart (Grant, 2007). However, the research supports the proposition that firms without residual claimants can get the labor donations easier, because in his case motivated agents are less concerned with self-interested behavior of the principal (Francois and Vlassopoulos, 2008, Besley and Ghatak, 2005). Opportunistic behavior of the principal can be more often expected in the for-profit, than non-profit settings, where principals are residual claimants. That is why in order to receive labor donations private firms should develop a reputation for allowing their workers' extra effort contribution to have an impact not appropriated by the principal (Francois and Vlassopoulos, 2008).

Fehrler and Kosfeld (2014) also ask themselves whether mission can motivate extra effort and test it in the laboratory labor market experiment. They allow their subjects to choose mission they are working for and compare their effort level to the control group where subjects can donate to a randomly chosen anonymous student. A special feature of their design is that half of the participants play the role of the employee, while another half plays the role of employer. Employer can offer contracts with fixed

wages and piece rate, which allows researchers to test the prediction of different contract choices. Employee chooses the level of effort, which determines pay-offs and donations.

As a result of the experiment researchers come to the conclusion that workers do not provide higher effort in the mission choice treatment than in the control treatment. This means that employers cannot save on monetary incentives in the mission treatment. In fact contracts they offered were not different from those in the control treatment.

In an attempt to explain the contradictions between their results and empirical and theoretical evidence of motivated workers in mission-oriented organization researchers once again turn to the selection process and argue that there is probably only a subgroup of workers who can be altogether motivated by mission. Paying lower than market wages can drive the self-selection of employees into such organisation.

In order to test this proposition Fehrler and Kosfeld (2014) conduct a second experiment of similar nature but combining the between- and within-subject design. As a result researchers come to the conclusion that approximately one third of participants chooses the contract where they can donate to the NGO of their choice. This group provides significantly higher effort than other subjects, thus, allowing the researchers to confirm their proposition that there is a motivated sub-group of workers which self-select into the mission oriented jobs.

Thus, it is the research in public (welfare) economics where we can find evidence that in the occupations, which are traditionally considered as meaningful (medicine, art, science, pedagogy) individuals derive pleasure from promoting noble goals and agree to lower reservation wages. However, economists believe that it happens due to the fact that some employees are predisposed to having altruistic preferences towards their customers. Such employees provide labor donations in form of extra effort without direct monetary incentives. Further the literature suggests that the specifics of the public sector (its mission-orientation) is what attracts such employees and leads to the situation of self-selection.

This conclusion is a logical continuation of the classical economic idea of seeing social preferences independent of the context. In other words being predisposed to exhibit certain level of social preferences people reveal their preferences by means of selecting into the particular field. Although it can be true that people tend to self-select into the specific areas, which match their primary job motivation, we believe that attributing such qualities only to some professions "can limit our understanding of meaning-in-labor" (Ariely et al, 2008, p.674). We argue that not (only) the mission, but the potential impact of the job on recipients makes some occupations more prone to motivate prosocial behavior. This is related to the nature of social services on one hand, and to the mechanisms, which we described above, including purely altruistic preferences of employees or desire for social recognition (Francois and Vlassopoulos, 2008)

In fact in the mission-oriented public sector the employees are often given the opportunity to have a chance of direct contact with recipient, who is perceived as needy, and to enjoy greater visibility of their social activity, which leads to their willingness to exert higher effort on behalf of their companies in form of labor donations.

We argue that under some constraints a much wider range of companies have an opportunity to provide their employees with comparable conditions. Though self-selection hypothesis has its support, we move the attention to the managerial interventions related to the employee's on the job performance. Also if we assume the heterogeneity of employees, as for example Ben Ner (2013) does, we face the need to motivate the existing work force. By means of changing the contextual parameters of the jobs the management can influence the perceived job impact and thus motivate prosocial behavior for the employees of different occupational fields.

2.5.3. Evidence from experimental research in management

While experimental laboratory research is very popular in economics nowadays, it is criticized by many for its lack of "realness" and problems with external validity. That is why it is very interesting to observe that the conclusions similar to the ones achieved in economics were also reported in the experimental and field research in management and psychology.

As management and psychology literature uses terms very different from those accepted in the economic field, and methods of the research are often seen with suspicion by economists, there is comparatively little communication between the fields. As a result, while talking about very similar phenomena, researchers frequently do not understand each other.[20]

It is also very important to see that in this stream of research the antecedents and consequences of prosocial behavior have been often studied in a much applied way. It means that context has been extensively introduced into experiments, so that relationships at work (including employee-customer relationships) have been already set in focus by this research.

For example, job performance being one of the most theoretically and practically important problems in organizational research, is tackled differently in economic and management research. In particular management researchers often take a constructivist approach (versus. deterministic approach popular in economics). In practical terms it means that in modern managerial literature many scholars have long recognized that performance depends not only on the objective characteristics of the jobs, but that it strongly depends on how the employees perceive their jobs. Thus, in order to increase the performance employers should change the perception of their job by the employees.

---

[20] Many concepts (e.g. a concept of job meaningfulness) have very wide definitions and so much personal components that they can be hardly operationalized. Hence, it is difficult to quantify and incorporate such concepts into quantitative economic analysis. That is the reason why, according to Ariely (2008) "identity, pride, and meaning are all left out from standard models of labor supply" and the literature on their impact on behavior is very scarce, especially within economics.

The existence of social preferences in the decision makers' utility functions received attention in the management literature in the form of discussion of prosocial motivation (e.g. Batson, 2008; Grant and Berg, 2010). In this literature prosocial motivation has been defined as the desire to have a positive impact on other people. At the same time the stress has been made on the fact, that prosocial motivation should not be equated with altruism, since "it refers to a concern for others, not to a concern for others at the expense of self-interest" (Grant and Berry, 2011, p.77).

This stream of research combines the traditional job design model of Hackman and Oldham (1976) with the ideas from social information processing theory. In particular Grant (2007) expanded the job characteristics model of Hackman and Oldman (1976) by including the notion of relational architecture of the jobs, which in his view can directly influence the motivation to make a prosocial difference. One of the options to do it is through variation of the perception of task significance - judgments that one's job has a positive impact on other people.

While job design researchers (Hackman and Oldham, 1976) argued that task significance was an objective characteristic of the job itself, they suggested to redesign the tasks in order to make them be perceived as more significant. Social information processing researchers, on the other hand, think that task significance is a subjective socially constructed judgment, so in order to increase the perceived job significance they propose to use social clues that alter employee's perception (Grant, 2008). Importantly both admit that the perception of task significance can be altered.

Classical mechanisms linking task significance and performance include the parameter of meaningfulness of work – that is purposefulness and value of the work (Hackman and Oldham, 1976).[21] Grant (2008) suggested there are also other mechanisms. He believes that relational mechanisms (processes that influence employee's connection to other people) play a big role in the process. According to the social information processing perspective, employees want to experience their actions as related and connected to other people. And task significance is doing exactly this, signalling to employees that their actions influence other people, who benefit from their job. It allows the employee to move from "an abstract, intellectual awareness of opportunities into a concrete emotionally driven understanding that one's personal actions can make a difference" (Grant, 2008, p.110).[22]

The model offered by Grant (2007) claims, that two specific attributes of relational job architecture – namely **job impact** and **contact with beneficiaries** – can influence the workers' behavior and consequently the outcomes of her actions. Contact with beneficiaries is defined as the degree to which a job is relationally structured to provide opportunities to employees to interact with the beneficiaries. It is believed to positively influence the affective commitment to beneficiaries.

---

[21] Meaningfulness has been a popular notion in psychology and management and it has been shown to be the key mediator of all five core job characteristics addressed in the JCM of Hackman and Oldham (1976): skill variety, task identity, task significance, autonomy, and feedback (Thakor and Joshi, 2005).
[22] And emotions motivate to actions as in Small and Loewenstein, 2003

Under job impact Grant (2007) understands the degree to which a job provides opportunities for employees to affect the lives of beneficiaries.[23] It allows employees to feel how their own actions improve the welfare of others (while task significance shows how job can provide opportunities to benefit others). The idea is to make it salient that others depend on employee's effort (Grant, 2008).

Grant (2008) conducted a set of field experiments aiming at examining the effect of task significance on job performance and mechanisms moderating this link. First with fundraisers (Grant, 2007), who were soliciting alumni donations to the university. The design of the experiment included the opportunity for the callers to communicate within 10 minutes with the student, who benefited from their work by getting the scholarship. The student explained how the scholarship made a difference in his life. As a result one month after the treatment had been applied, callers who talked with the student spent almost twice more time on the phone and twice more donations compared to the control group (who did not communicate with the student). There were, however, individual differences, which moderated such effect. In this study Grant manipulated two parameters: perceived task significance (information on how the other benefited from the work) and contact with the beneficiary (before callers did not have such contact).

Further in order to separate the effects of task significance and contact with beneficiary Grant (2008) removed the contact with beneficiary and made an attempt to understand the mechanisms linking task significance and performance. The experiments with callers (who this time were provided with stories about the beneficiary of their work and not with personal contact) showed that the treatment group more than doubled the amount raised one month after the intervention.

Another experiment with lifeguards at a community recreation centre, who read four stories about other lifeguards rescuing swimmers increased job dedication and helping behavior, as showed by the analysis of hours of volunteering before and after intervention, and questionnaires from their supervisors and themselves. And finally, the experiment with newcomer callers showed that "callers with low levels of consciousness and strong prosocial values were most responsive to task significance" (p.118)

Similar understanding of job meaningfulness can be found in the research conducted by Thakor and Joshi (2005). The researchers used job characteristics model of Hackman and Oldham (1976) to argue that meaningfulness of the work is one of the most important mechanisms for motivating sales employees to engage customer oriented (versus. sales oriented) behavior.[24] Customer orientation is conceptualized as focus on uncovering and satisfying the customer's long-term needs by listening and working with customers. While sales-orientation suggests the satisfaction of customers' apparent (articulated) needs, customer-oriented employee should go much further in order to discover deeper

---

[23] The examples of jobs with high parameters of job impact are surgeons, automotive design engineers and lifeguards, while cashiers and gasoline station attendants have low impact on beneficiaries.
[24] The other two are organizational identification and pay satisfaction.

needs and preferences of the customer. Such behavior requires greater effort on the part of employee. In order to motivate employees to provide such extra effort Thakor and Joshi (2005) offer such mechanisms as experienced meaningfulness and organizational identifications, which together foster the feeling of accomplishment. Thakor and Joshi (2005) propose the following chain: if the perceived importance becomes more apparent, employee experiences greater meaningfulness, which in its turn raises belief that engaging into activity will bring the feeling of accomplishment. Such feeling is a part of intrinsic motivation and triggers higher effort. They believe that their results highlight the managerial importance of ensuring that salespeople find their job meaningful.

Another mechanisms for mediating the task significance – performance link, according to Grant (2008) is perceived social worth – the degree to which employees feel that their contributions are valued by other people (or believe they are appreciated by others). According to Grant (2008) "pursuit of social worth is a basic human motivation". We believe that this aspect of prosocial motivation is related to the warm-glow and norm-based sources of prosocial motivation. Consequently, they can be strengthened by making the actions of the employee observable, or make her visible and identifiable to the customer and wider community.

Also Grant (2008) argues that a set of personal traits can mediate the link, in particular conscientiousness and prosocial values. Task significance is more likely to increase the job performance of the less conscientious employees, since for more conscientious ones the good performance is a reward in itself. The same intervention is also more likely to increase the performance of the employees with strong prosocial values, seeing the importance of promotion the welfare of others as guiding principle in life.

In our opinion such complex view of prosocial motivation leads to the understanding that a wide range of individuals can actually be motivated by the desire to make a prosocial difference.

Thus, the research in management, especially the agenda proposed by Grant and colleagues (2007, 2008, 2010) shows that prosocial behavior can be promoted by contextual factors, which can be altered by managerial interventions. In particular employers can help the employees to see their work as meaningful and high in impact by means of introducing relational job design. Contact with customers is one of the most important elements of such design. Obviously, there are certain factors mediating the link between the contact with the customer and the effort (e.g. characteristics of the job, individual traits, etc.), but in general researchers agree that "any job can be experienced as contributing to others' welfare" (Colby et al, 2001 cited in Grant, 2008).

III. Hypotheses

3.1. From research question to hypotheses

While prosocial motivation manifests in the relations between supervisors and subordinates, among colleagues, as well as with wider community, we are interested in the specific type of prosocial motivation which addresses the relations in the employee-customer dyad. In particular, based on the analysis of the literature, we argue that the information about beneficiary (customer) can influence the prosocial motivation of the giver (employee) by means of strengthening the empathic concern, while in case when the customer is provided with information about the employee, the latter can have a stronger motivation to behave prosocially under the pressure of image concerns.

In light of the above, the general propositions can be stated in the following way:

Proposition 1: The employees are willing to make higher labor donations[25] when they perform a task for the identified customer (by means of the name, photo, and personal information), than for an unidentified (abstract, general) customer.

Proposition 2: The employees are willing to make higher labor donations when they perform a task for the customer who can identify them (by means of the name, photo, and personal information), than for the customer unaware of the employee's identity.

In the proposition 1 we employ the idea that providing information about a selected individual customer leads to the reduction of social distance and thus to higher chances for the employee to experience empathic concern. Empathy (or empathic concern) related to the customer leads to the motivation to perform a better job on behalf of the customer, and thus higher effort provided by the employee.

In the proposition 2 we argue that by making the employee aware of the fact that her identity is to be opened to the customer, we increase the observability of employee's actions and her level of effort, thus making her image concerns more salient. Image concerns motivate higher effort of the employee.

While in both cases we talk about the extra effort provided by the employee, the sources or motives in these two cases are apparently different. If empathic concern moves the focus towards the needs of the customer, the image concerns return it to oneself. We hypothesize that such inherent difference in motives influences not only (or not necessarily) the magnitude of effort, but its direction. Such distinction is especially important in management research as some tools promoting apparently other-regarding behavior can have adverse effect on output. That is the reason why we go one step further in our hypothesizing and relate the above mechanisms (customer and employee identity disclosure) not

---

[25] Labor donations are reflected in the effort not compensated by monetary rewards.

to the magnitude, but to the direction of effort. Using the context of multiattribute tasks allows us to come up with testable hypothesis.

3.2. Incorporating the complexity: trade-offs in the multiattribute task

In most principal-agent relationships, the principal has to induce the agent to engage in several tasks simultaneously: for example a worker who should produce certain amount of output that is probably easily measurable, but who also should probably assure high quality of output and maintenance of the equipment. Also a school teacher should teach some basic skills which can be controlled by means of the test, but she should also stimulate students' creativity, teach social skills, which is much more difficult to do and also to measure (Fehr and Schmidt, 2004).

In economics incentive provision in multitask principal-agent model is described in the seminal paper of Holmstrom and Milgrom (1987). There is an interaction effect between giving incentives for performing one task and agent's incentives for performing other tasks. In particular, if the tasks are compliments and the principal wants that the agent gets involved in all tasks, it is necessary to reduce the incentives on the task that is easy to measure (compared to the single task setting) and introduce high-powered incentives in the task which is difficult to measure. Otherwise the employee will pay more attention to the easier measurable task disregarding other tasks.

Importantly Fehr and Schmidt (2004) admit that such distortions produced by explicit incentives in multitask environment may be the reason for introduction of implicit incentives in form of voluntary bonus payments. Bonus is paid based on the subjective evaluation of the whole result by the principal (and cannot be verified in court).

Fehr and Schmidt (2004) conduct an experiment where they allow the principal to decide which type of contract she wants to offer to the agent. They come to the conclusion that bonus contracts are preferred to fixed-wage and piece rate. They explain it using their theory of inequity aversion (1999) stating that 60 percent of people are selfish and 40 percent have equity concerns. They claim that it is more profitable for selfish agent to work under bonus contract than for the inequity adverse one: the latter loses more if the bonus is not paid. At the same time inequity averse player works better under generous fixed wage, as selfish one still provides effort only on verifiable task.

While such direction of scientific reasoning represents an attempt to use the monetary incentives to motivate balanced approach of the agent towards multiple tasks, present research explores the potential of non-monetary incentives for the same purpose. We also assume that those tasks where performance is difficult to monitor or to measure are also the ones, which are more difficult to perform, because they do not have a clear script and employees need to exert effort to come up with the script for themselves.

Even more important is that in many cases there are no clear quantifiable output requirements providing the agent with ready-made prescriptions for quantity / quality trade-offs or such requirements are so vague, that under the environmental uncertainty the agent can always claim that the results are caused by external factors.

So beyond the possibility to distinguish between two motivational mechanisms, the choice of multiattribute task has the following reasons:

1. Multiattribute tasks are most common in modern world and the problem of balancing the effort of the employee, especially in the presence of uncertainty and lack of universal requirements from the management, is important.

2. In multiattribute tasks we can observe attributes where measurement is easier and others where monitoring requires more effort. Even in the absence of direct monetary punishment attention to such tasks can be different. At the same time some attributes can be more important for the customer than others. Thus, paying more attention to the highly visible attributes will signal the mechanism of image concerns playing a role. Paying more attention to the customer-focused parameters will signal empathy working.

We chose a multiattribute task where direction of effort represents a trade-off between parameters. Our question is not as much about the absolute magnitude of effort, as about the relative trade-offs that employees make when exposed to customer- or self-identification.

3.3. Final hypotheses

While both empathy and image concerns can result in the increased amount of effort, we argue that in a multiattribute task the direction of the effort will be different depending on the mechanism staying behind the motivation.

Thus, our main research questions are the following:

- Does opening customer's or employee's identity change the focus of effort in multiattribute tasks?
- Does opening customer's identity motivate a different focus of effort than opening the identity of the employee?

These research questions bring us to the following line of argument: opening customer's identity leads to reducing social distance which in its turn motivates empathic concern. Since empathic concern relates to the genuine focus on the needs of the other person, we argue that the employees exposed to the customer's identity will be more willing to meet customer's needs, thus focusing on the parameters of output directly benefiting the customer.

At the same time making the employee's identity known to the customer increases the observability of the employee's actions and makes image concerns more prominent. We believe that this results in the employee's desire to "show" that she exerts effort in order to benefit the customer (which is not necessarily equal to benefiting customer in reality). In other words the employee will focus on parameters of output signalling the effort.

If attributes of the task directly benefiting customer and signalling effort are one and the same then the only difference we can observe is in the magnitude of effort. If, though, these attributes are different, then opening employee's identity can have a detrimental effect on the output.[26]

As a result of previous discussion we come up with the following hypotheses:

Hypothesis 1. In multiattribute tasks making the individual customer identified by the employee moves the distribution of effort towards achieving the customer-benefiting trade-off among different output parameters.

Hypothesis 2. In multiattribute tasks making the employee identifiable to the customer moves the distribution of the effort towards achieving the trade-off among different output parameters perceived as able to signal employee's effort.

There are also some additional questions, which we are going to test within our research. In particular we are also interested in the possible results of combining the effect of customer and employee identity disclosure. We suggest that such combination provides the employee with wider spectrum of motives and can lead to a simultaneous focus on all parameters of output. The effect of customer identification will moderate the negative consequences of self-image focus.

Another issue is related to combining the behavioral and dispositional view of motivation, so called trait versus state perspectives. We can suggest that higher empathic ability (empathic trait) will always move focus towards the customer-benefiting parameters of output. In other words, if customer identity disclosure motivates empathy-based behavior, the effect will be stronger for the employees with higher initial empathic ability (empathic trait). At the same time in case when employee's identity is opened and image concerns are governing the behavior, higher empathic ability will move focus away from image-benefiting parameters of output towards customer-benefiting ones.

After we have introduced the experimental design, we will specify the hypotheses directly linking them to output parameters of the particular multiattribute task used within present research.

---

[26] Given that customer satisfaction is an important goal for the employer.

IV.   Method of Research

4.1. Experimental method

4.1.1.   Reasons behind the choice of the method

In order to support or reject our hypotheses we should be able to observe (statistically) the difference between effort levels, measured in terms of quantity and quality of outcomes for the employees having no information about a particular customer they are working for (general customer) and not exposed to the customer observation and those influenced by these treatment variables.

As was mentioned before, much of the support for the behavioral phenomena comes nowadays from the field of experimental economics. This extensive use of experimental approach is caused by the fact that experiments not only allow to directly observe the behavioral phenomena, but it also provide an opportunity to study causal relationships (Antonides, 1991).

Talking about experiments we should distinguish between laboratory and field experiments. Field experiments are characterized by the real world settings and the opportunities they provide to draw the conclusions related to the processes naturally occurring in the field. Laboratory processes are much simpler, but they enable the researchers, by keeping all other factors controlled, measure the impact of treatment effects on the behavior. The level of control achieved in the laboratory is difficult (or even impossible) to obtain in the field.

Our research agenda required settings close to the ones in the field experiments but the challenge with using this method is twofold: first in natural environment it is difficult to argue that the difference in observed behavior can be attributed solely to the treatment effect (absence of direct control on the part of the experimenter) and second – opportunities for finding and acquiring permission to operate in such environment is are rather limited.

At the same time traditional laboratory experiments, as well as the general reasoning, intentionally avoid the influence of the social context, since by introducing social context the experimenter loses much of the control. According to Huettel and Kranton (2012) most of the experiments are conducted at the individual level and anonymous interaction level, which means they are largely devoid of social context. At the individual level there is no interaction and action have only personal consequences. The assumptions in such settings belong to the field of traditional economics of self-interested anonymous individuals. At the level of anonymous interaction partners intentionally made anonymous, and any social cues and context are stripped away. There is interaction, but without context, like in behavioural economics, social psychology and social neuroscience.

Our research questions put context into the centre of discussion, which requires usage of different types of experimental design as described above. Huettel and Kranton (2012) describe another two

types (levels) of experiments implying increasingly higher share of social context. Both of them involve relationships between subjects in the lab, but the "generic social context level" does not address the issue of social identities or groups. In "identity level" designs experimenter either induces identity level or identities outside the lab are incorporated into the experimental design.

Following the similar logic we moved in our experimental design from the classical experimental settings and enriched them with context.

Our research question can be roughly fit into the design of one of the most researched games in experimental economics, namely dictator game (further DG). This is the most popular game employed in the research on altruistic behavior. The idea of the dictator game is that the proposer, who has a specific endowment (usually \$10 or €10), has an option of sharing of this endowment with the responder. The responder has no say over the distribution, which, from traditional perspective, should mean that the "classical" proposer keeps all the money. The first DG has been conducted by Kahneman, Knetsch, and Thaler in 1986 and gave subjects not real, but hypothetical choices with even or particular uneven split of endowment. According to Camerer (2003) usually more than 60 percent of givers transfer positive amount of money and the mean transfer is about 20 percent of the endowment.

In the traditional framework the dictator game is a classical economic game (response game) where utility is represented by monetary assets. At the same time it is effort and not money, which is shared by employee in her relationships with both employer and customers. In fact, altruistic motivation can lessen the disutility of effort, but can have a more neutral influence on the utility of money, i.e. one is more willing to exert higher effort, but not willing to give away money (Brüggen and Strobel, 2007). This is another reason for putting the decision about the individual level of effort exerted for the particular task in the centre of experimental design. In particular, we believe that real effort tasks should be in the core of the experiment.

Another aspect of the design is related to the meaningfulness of the task for the participant. In the majority of existing experimental designs tasks were deliberately deprived of meaning: subjects crack walnuts, perform cognitively demanding tasks on the computer (two-variable optimization problems) and solve mazes (Fahr and Irlenbusch, 2000; van Dijk, Van Winden et al, 2000; Gneezy, 2003).

By choosing the meaningless tasks experimenters try to exclude the possibility of intrinsic motivation (Brüggen and Strobel, 2007), but the downside of this approach is that this may prevent the subjects from behaving like in a real job environment. In our case employee should be aware that her effort leads to the outcome (product), which brings direct utility to the beneficiary In particular if we talk about the distinction between quantitative and qualitative characteristics of output, then in order for the employee to provide high-quality output, she should understand, how the outcome of her effort is

used and what characteristics of it are important (why quality is important). It is hardly possible with artificial meaningless tasks, where employees cannot really identify with the goal of the task.

Also the experiment should guarantee that the incomplete contracting condition is met: employee receives fixed compensation for her effort unrelated to its amount and is free to determine the amount of effort she wants to exert. At the same time employee should have a one-shot relationship with the beneficiary (in order to avoid reputation effect).

Thus, in our case the employee can be seen as the dictator, who has the right and ability to decide about the amount of effort she is willing to provide. The customer has basically no say about this amount of effort. The conditions of incomplete contracting, with special stress on non-contractibility and hence, absence of material rewards for additional effort should build the necessary context for our discussion. So the question can be translated into economic dimension as the one about the willingness of the dictator to give up some of her utility (in our case in form of her effort) in order to increase the utility of the beneficiary (in form of final product or service). It is very important for our analysis to distinguish between the effort of the employee motivated by genuine concern about customer's needs and effort caused by the image-related concerns.

As it has already been shown above, contemporary research in economics provides a new foundation for our study, as it assumes that the discussion of prosocial motivation should encompass the contextual dimension. In particular, contrary to the traditional unconditional view of pro-social behavior, many researchers nowadays view the information provided to and about the parties as one of the important contextual factors mediating prosocial behavior. Experiments testing the influence of information on behavior often focus on communication between the parties.

At the same time, using communication as in terms of contextual dimension, we should remember that its impact strongly depends on its direction (e.g. from giver to the receiver versus. from the receiver to the giver), its richness (e.g. verbal, non-verbal, with or without the identification) and content (e.g. what is allowed to be communicated). There is a long discussion in the literature concerning the effects of direct face-to-face communications. Although it is the most wide-spread type of communication in everyday life, economists argue that introducing this type of communication into the experimental designs can limit the researcher's ability to make inferences about the direct causes of behavior. Such communication is a process going in both directions, with almost simultaneous exchange of information, both in verbal and non-verbal form. The final behavior can be a product influenced by a whole range of factors, which can be hardly distinguishable from one another.

As for the direction of communication, it is important, as previous research has shown, to distinguish between mechanisms operating in case of one-sided giver-received versus. receive-giver communication. Also in cases of mutual communication the sequence will play a particular role. We use the one sided communication always starting from the information provided by the receiver and

later asking (in some treatments) for the information from the giver. This allows us to exclude the possible influence of the communication sequence as an intervening factor.

We mostly use the written verbal communication as a tool isolating the influence of non-verbal clues. However, knowing that non-verbal clues can be responsible the big part of information transmitted, we will mitigate the lack of informational content by adding the visual information for information stream going from receiver to the giver. In particular we provide the employees with the photo of the customer, thus not only lifting the condition of anonymity, but also increasing emotional appeal of the message. Arguing that empathy is one the mechanisms motivating socially-oriented behavior we believe that for this mechanism to be at work, it is essential to show the real beneficiaries employees can identify with and facial images can be very important in this respect.

We also believe that it is quite important to keep the content of communication largely constant with differences only in the particular parts triggering treatment effects. Our aim is to see if the slight medication of the (same) "message" can contribute to the change in behavior.

That is why we narrow down the scope of information exchanges in our experimental setting by focusing only on written verbal information with strictly controlled context. Our goal is to see if such contextual factor as presence of personal information can have an impact on prosocial behavior. In order to achieve this goal we should eliminate the possibility that strategic concerns (like reputation building) or expectations (e.g. in form of reciprocity) can influence the behavior of givers (employees) towards recipients (customers). Our task, thus, is to avoid direct communication, but introduce some other limited forms of communication. We believe that for the mechanism of empathy to be at work, it is essential to "show" the real beneficiaries whom people can identify with.

We argue that lifting the condition of anonymity by means of providing the subjects with identity information about the other party and their facial images (photo) is enough to lead to higher effort on the part of employee. In order to test this influence we propose to provide the giver (employee) with the information about the recipient (client) and investigate the influence of such manipulation on the prosocial behavior of the former. At the same time we suggest that informing the employee that her identity is to be opened to the customer can also result in higher effort of employee. In order to test this hypothesis the recipient should be provided with information about giver, which giver should be informed about. Combining two directions of informational supply allows testing the influence of two-sided information exchange on the prosocial behavior of the giver.

Following the discussion in the Literature review and operationalizing our hypotheses we introduce the following treatment variables:

- Non-anonymity of the beneficiary – giver is provided with the information about particular beneficiary of her effort

- Non-anonymity of the giver – information about giver is made open to the beneficiary.[27]

4.2.2. The experiment: organisational issues

Experimental design is based on the idea of real effort experiment, where participants are supposed to transcribe a recording of a lecture into a text. The experiment was conducted at the University of Vienna in collaboration with the non-profit project of the University of Vienna "Mitschriften Börse", which has an aim of providing the students with hearing impairments with studying materials required for completion of particular courses. Within the framework of this project interested students are provided with audio recorders for recording the lectures they need. These recordings - in order to be used by the intended beneficiaries (students with hearing impairments) - should be transcribed and formatted to become a clear readable text. Such transcription work has been used as a major task in our experiment.

The specific of the task allows us to distinguish between the quantitative and qualitative parameters of the desired outcome. From the quantitative perspective the more recording time is transcribed and the more words are typed, the better. On the other hand, qualitative imperfections - mistakes and typos, lack of structure (paragraphs, numbering, etc.) - can hugely diminish the value of the transcriptions and even make the text unusable for the beneficiary. That is why quality of the outcome plays also a very important role.

However, it is quite costly to control the quality for the "employer". Moreover in this task the quality requirements can hardly be made in the one-and-for-all mode: although spelling accuracy (further addressed as accuracy of the transcript) can be universally agreed upon, the structuring of the text and performing general formatting in order to make it more readable requires some "creative" effort from the participants. What is also important is that maximization goals for both quantitative and qualitative parameters of output require specific trade-off point, which can hardly be set by the management and is in fact determined by the employee. The employees need to put themselves into the beneficiary's shoes and end up with the output (transcript) which they themselves would find appropriate.

In other words if there is no certainty for the principal regarding the necessary standards of output (when taken into account the resource constraints) it could be important that employee herself takes the decisions about the right combination of attributes. For example the principal can be willing to maximize all parameters of output, but it can be impossible due to the time constraints.

Moreover there is a large number of tasks where output can be hardly measured and sanctions are difficult due to the high level of uncertainty.

---

[27] It is important to explore the observability of the giver to the beneficiary, because she (the beneficiary) is the one who can determine the satisfaction of the quality standards, which means that quality can be addressed better if givers identity is opened for the beneficiary.

We introduce all our subjects to the condition where they work in the frame of prosocial mission. However, we do not allow any self-selection. That is why we have a heterogeneous group of participants, some of them probably more mission-oriented than the others.

Since we do not have an aim of testing the self-selection hypothesis, we intentionally from such heterogeneous groups. If we suggest that due to the randomization we have a similar level of heterogeneity in every treatment group, then the differences between groups would be addressed to effects other than mission orientation.

We do not test directly the relationship between personal traits and working behavior, although we use some measurement instruments for assessing individual level of dispositional empathy and agreement with norms of helping and social value orientation.

While designing the experiment we had several options for setting performance measures and consequently linking them to the payment:

1. Participants are paid piece-rate for every minute of the recording they have transcribed. Such design does not provide us with incomplete contracting condition, which is necessary for testing the mechanisms of prosocial behavior.

2. Participants should transcribe a fixed length of recording (e.g. 20 minutes). Payment is fixed for the pre-set amount of recording and is not linked to the quality of the outcome. Problem with this design is that depending on the abilities, participants can need different amount of time for completing the task. For some the task can even be too difficult to complete. We could require a shorter recording piece (the shorter the required length of the recording - the easier it is to complete the task), but from the "realness" perspective, possibility of observing customer-driven motivation in case of a task with considerably little impact is rather small (e.g. 10 minutes of lecture transcribed cannot bring much to the beneficiary). In general such approach can make the whole experiment overly dependent on the task proficiency of participants.

3. Participants should transcribe a recording within a fixed amount of time (e.g. 1 hour). Payment is fixed for the pre-set amount of time and it is not linked to either the quantity or the quality of the outcome. The performance measure is formulated as "do as much as you can". The experiment has a fixed end-point in time, which is easier to organize. In this design participants have even less motivation to exert effort for monetary reasons, which allows more credibility in explaining the effort exerted by social motive.

We selected the third option for our experimental design and determined two sets of performance measures in our experiment:

1.Quantitative: how much in minutes of the recording and in words typed a participant transcribed within the pre-set amount of time.

2. Qualitative: how precise was her transcript (whether she typed every word she heard on the tape or not), how many mistakes and typos she made, and finally if proper paragraphing, numbering has been provided[28].

One of the very important questions in this experimental design is the necessity to control for the initial difference in task proficiency between participants. We control for this difference by means of introducing an extra phase in the experiment - control time (pre-treatment) where participants work within 10 minutes without being informed about the purpose of their effort. By measuring the effort of the employees within this time we receive a benchmark for the further analysis both within and between treatment groups. Having this control time incorporated into the experimental design allows use of the statistical tools for related design samples, as well as employing regression analysis with parameters achieved by employees in the control time used as control variables.

Thus, the experimental procedure was split into three parts (Figure 1), each characterized by its own set of output variables. In the control time, participants worked without knowing the purpose of their effort. In the main part of the experiment, where participants are paid for their effort, they should have been motivated to provide higher quantity of outcome, while keeping to some quality standard. The additional part was designed in such a way, that participants were not paid for their effort and had complete freedom of decision regarding the very decision to participate in it. At the same time this part was explicitly aimed at allowing the employees to improve their output in terms of quality.

Figure 1

Phases of the experiment

	Introduction of Treatments		
Pre Qs	Control time (10 min)	Main time (50 min)	Additional time (up to 30 min)	Post Qs
	t0	t1	t3	
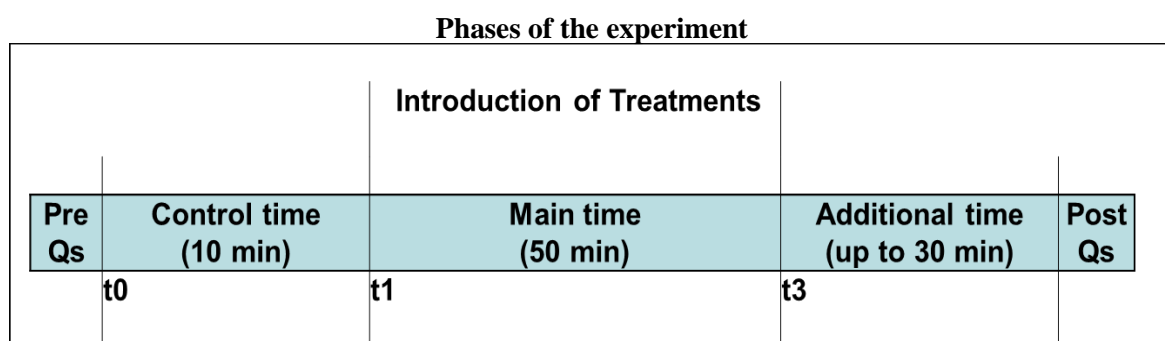
Each participant was provided with personal earphones connected to a laboratory PC. We used two types of software: one for playing the recording and another for typing the text. For playing the recording we used VLC Mediaplayer with the function of back play, slower and faster play operated by using key-combinations on the keyboard, while the player itself is in inactive mode. For the typing

---

[28] Quality of formatting is not included in the main analysis, but contributes to the analysis of employee's behavior in the additional time.

we used MS Word. We are aware that there are some professional tools used for transcribing purposes, but we believe that they could have some drawbacks in the context of our experiment: considerable time and effort is required to master them, and the very process of learning can be seen as either intrinsically motivating or on the contrary demotivating for different types of participants. We believe that the intrinsic value of learning can undesirably influence the results. Also time constraints determined the necessity of shorter learning time. That is why we decided to use the software familiar to all participants and not requiring much learning and MS Word was the best solution. VLC Mediaplayer is also a software which is very user-friendly as it can be operated by a few pre-set combinations of keys while having the program itself in inactive mode.

4.2.3. Experimental procedures

The experimental sessions were organized as follows. Students, who responded to the announcements (Appendix 1) were invited for participation according to their selected dates and times. Participants were randomly assigned to the working stations, which in their turn were randomly associated with the treatments. In the beginning participants filled in two questionnaires – Social Value Orientation slider and Empathy trait questionnaire (Appendix 2)[29].

After that participants were provided with the instructions stating that they should perform a transcription task (Appendixes 3 and 4). In particular the participants were informed that they would be translating the recording of the lecture. They had to transcribe as much recording as possible and transcript had to be made word-for-word. The participants were instructed on how to use a Media Player which was playing the recording of the lecture. The Media player was already open on their PCs. They had to type the text into the MS Word file also already open on their PCs. The combination of Media Player and standard application MS Word was chosen intentionally in order to simplify the transcription process and avoid extended learning process. Participants needed only four key combinations to play/pause the recording, as well as move it backwards or forwards. We asked participants to try each combination and made sure that each of them grasped the concept of using these combinations. After they had tried the key combinations and had their questions answered (if any) participants were instructed to start transcribing and had to work for 10 minutes. When this time was over we distributed the second part of the instructions containing treatment information (Appendix 5). In particular all the groups received the following information (Figure 2).[30]

---

[29] Will be described in more detail in a separate section below.
[30] The full version of Instructions is provided in the Appendix 5.

Figure 2

Selected parts from the Instructions provided in the main time
(part identical for all groups)

Today you are employed to do a real job task - a transcription of a recording of a lecture.

This study is conducted by the Chair of Organization and Planning in collaboration with the project *Mitschriften Börse* organized at Vienna University. The aim of *Mitschriften Börse* is to provide the students with disabilities better chances of access to higher education. In particular *Mitschriften Börse* assists the disabled students in getting learning materials necessary for completing the courses. There are students (e.g. deaf) who are unable to make lecture notes, which in many cases makes completion of the courses impossible.

To allow such students a better access to the learning materials, *Mitschriften Börse* provides them with the opportunity to get the transcripts of most important lectures. These lectures are first recorded and then have to be transcribed (turned into the typed text) in order to be used by disabled students.

While the above part was completely the same for all four treatment groups, further we provided participants with the additional information which differed in not more than one aspect from at least one of the groups (Figure 3). In particular the Group 2 differed from Group 1 only in the amount of the information provided about the customer, and Group 3 – only in the requirement of non-anonymity of the employee. Group 4 differed in both of these aspects from Group 1, but only in one of them from Groups 2 and 3.

Figure 3

Visual representation of differences in the information in the treatment groups
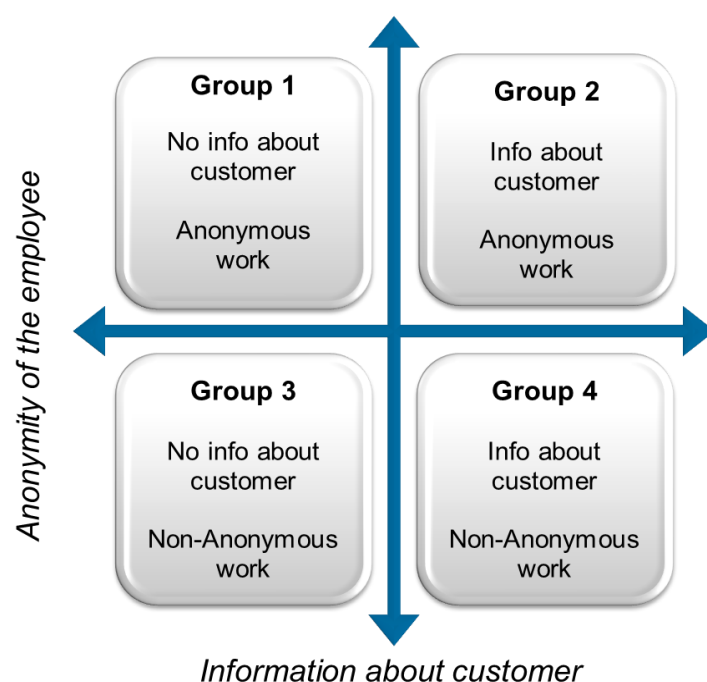
Below we show how these differences in the information were worded in the instructions (Table 1).

Table 1

Wording of instructions introducing differences in the amount of information about the customer and the amount of information provided by the employee

Differences in the amount of information about the customer	
Groups 1 and 3	**Groups 2 and 4**
Today you are employed to transcribe a recording of a lecture for the Mitschriften Börse.	Today you are employed to transcribe a recording of a lecture for Alexander Friedrich, who is one of the students registered within the Mitschriften Börse. Alexander is 28. He is a student of the Faculty of Sociology of Vienna University. This semester he has applied to the Mitschriften Börse asking for the transcripts of the lectures in Sociology. Due to the injury that he has suffered several years ago, he is unable to make the lecture notes himself.
Differences in the amount of information provided by the employee	
Group 1 and 2	**Groups 3 and4**
Please remember that your work is completely anonymous. Neither your name nor any other information about you will be revealed to Mitschriften Börse.	Please remember that your work is not anonymous. We ask you to write your first and second name at the end of your transcript!

All groups received the same instructions regarding the requirements for quantity and quality of output: they had to transcribe as much recording as possible given that they provide a word-for-word transcript.

After the main working time was over (50 minutes) participants were given a third part of the instructions, where we stressed the importance of quality for the final output (Appendix 6).

Together with the instruction were provided final questionnaires consisting of the situational Empathy test as well as number of questions regarding social norms, personal evaluation of the conducted work and some demographic characteristics of the participants (the full list of the variables derived from all the questionnaires is given in the Appendix 7).

After the participants had finished answering the questionnaires, they submitted them to the experimenter, were paid and left.

4.2. A closer look at the main output parameters

4.2.1. General description of parameters

Each part of the experiment was characterized by a common set of output variables. In order to be able to measure the level of output at different points in time during the experiment the requirement to save the transcript at particular points in time was included into the instructions (Appendix 4). This was to be done by using a specific icon (this icon was linked to the macros allowing saving document without overwriting the previous version and it was added by the experimenter to the icon panel on each working station). In order to minimize the probability of guessing for the purpose of this procedure we have informed participants that occasional saving of the word document prevents them from risk of losing the whole transcript in case of technical problems.[31]

Thus, at the end of each time slot we have a separate transcript, which we can also easily relate to the relevant piece of recording. This means that we can operate with two quantitative parameters of output, namely amount of words typed and amount of the recording transcribed (Figure 4). We divide each of these parameters to 10 (minutes) to come to the speed of transcribing in words (Speed_Words_10) and speed of transcribing in minutes of recording transcribed (Speed_Minutes_10). While first parameter describes how many words the participant typed in 1 minute, the second relates to the minutes of recording (lecture) she has transcribed in 1 minute.

Figure 4
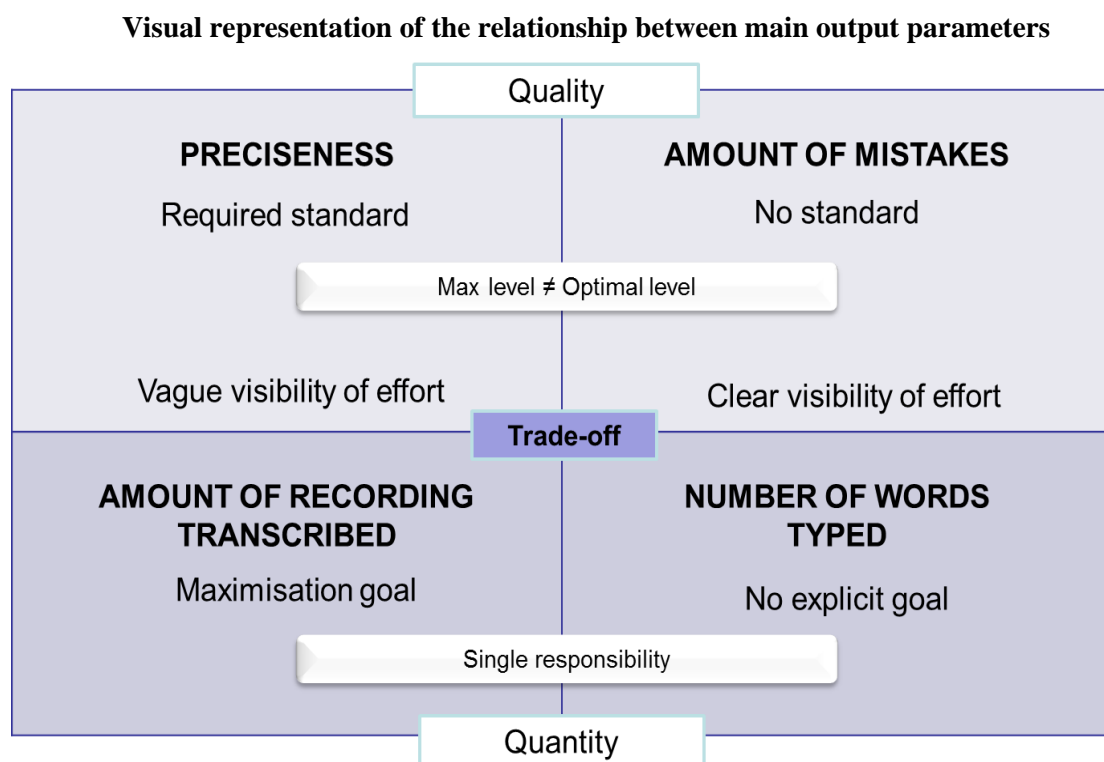
Visual representation of the relationship between main output parameters

---

On the qualitative side there are another two major output variables, namely preciseness of the transcript and number of mistakes. The first variable relates to the fact that despite explicit requirement to make word-for-word transcript, participants – in the absence of financial sanctions - in fact were free to decide on this issue. It is obvious that word-for-word transcription requires more time for the same amount of the recording, which means that by following this requirement less strictly participant could produce a less precise transcript, though covering more material from the recording. For example if two participants typed 100 words in one minute, for the person making word-for-word transcript this could mean transcribing only 20 seconds of the lecture, while for the one making less precise transcript 100 words could have been used for "summarizing" 2 minutes of the recorded lecture.

The second qualitative variable relates to the amount of misspellings in the transcript. In particular we calculate mistakes manually using Word function for checking the grammar and count each underlined word (word containing misspelling and/or word without first letter capitalization for nouns) or pair of words, if they do not contain space between them. This parameter, if other parameters kept constant, has the inverse relationship with the speed of typing.

4.2.2. Main output parameters in the context of the real-effort task

Thus, we have four main parameters of output available for analysis. Two of them are qualitative and two quantitative. In order to analyse and especially interpret them (also out of the context of the particular experimental settings) we need to understand what each of them generally means.

Before we analyse each parameter separately it is important to mention that our experimental design did not include any elements of monitoring resulting in monetary punishment. In other words none of the parameters were influencing the monetary remuneration, which was fixed at the level of 15 Euro for 1 hour of transcribing. Thus, participants could freely decide not only about their general level of effort, but also about the relative importance of different parameters of output. We argue that the treatments – or the information received by the participants within each treatment group – influenced the focus point, which could move from qualitative to quantitative parameters or the other way round.

Preciseness

Preciseness of the transcript - measured as number of words used by the participant to transcribe one minute of recorded lecture - is a parameter characterising the quality of output. In our settings this parameter is prescribed in the sense that we require a "word-for-word" transcript, which essentially means that each word of the lecturer should be typed by the participant. However, as described above, in the absence of monetary sanctions against violation of this requirement, the participants could freely decide on the importance of preciseness for them. It is also important to understand this parameter

from the point of view of the customer. In fact complete preciseness is not necessary for a "good" transcript. There are words and sometimes even sentences that do not carry any contextual sense.[32] At the same time typing these words / sentences is a time consuming activity which reduces the amount of recording which a participant can transcribe within certain amount of time. For example the precise transcript will mean more than 100 words for the minute of recording. However, only around 15% of the participants were providing this preciseness. Others were not transcribing "word-for-word". As we will see later from the results of our follow up study, higher preciseness is not necessarily associated with higher overall quality and such transcripts are not necessarily preferred over less precise transcripts having better quantitative parameters.

Preciseness can depend on the initial attitude of the participant towards its importance for a transcription purposes, but this importance can be revaluated, e.g. in case of direct requirement for achieving of certain standard (as in our case). At the same time changing the level of preciseness will have a direct impact on the speed of transcribing. That is why change of this qualitative parameter will have direct influence on the quantitative parameter. Hence, participants have to think about the trade-off.

In general context we can see the preciseness parameter as quality parameter of output which is though set by management, not necessarily linearly related to the perceived quality of output, especially as seen from the perspective of the customer.

Number of Mistakes

Number of mistakes was measured as the number of mistakes / misspellings in the transcript. We used the function of the MS Word for checking the grammar of the text and manually counted all underlined words.

In our settings this parameter is unrestricted and never mentioned within first and second parts of the instructions. It is only highlighted in the third part of the instructions distributed among all participants after the main time. In that part we draw the attention of the participants to the fact that quality is an important parameter for the customer and accuracy of typing and proper formatting may be seen as prerequisites of the good transcription.

Thus, in the main working time given the lack of monetary sanctions it was purely participant's decision to produce a transcript with higher or lower number of mistakes, and this was a decision which was not directly influenced by the instructions. We should remember that perfect accuracy comes at a cost of lower speed of transcribing, meaning that in an attempt to provide high quality of typed text one can transcribe less of the recording than if transcript was less perfect.

---

[32] Starting from words like "hmm", "na ja" etc. and ending with while sentences which do not have any direct relation to the topic of discussion.

In general context, the accuracy of typing can be seen as a quality parameter of output which does not have any standard set by management and it is not even mentioned in the job description. This parameter is fully dependent on the initial belief of the participant about the importance of accuracy for transcription. This parameter of output has however much clearer "common sense standard" in comparison with preciseness of the transcript. By saying this we mean that accuracy of typing can be easily recognized without listening to the recording and without reading the text carefully (as is necessary for preciseness). Short glance is enough to see the mistakes, like lack of capitalization for nouns (the recording and the corresponding transcripts were made in German) or lack of space between words. That the reason why we consider the accuracy of typing as a highly "visible" parameter of quality.

Speed of Transcribing

Speed of transcribing was measured as the amount of minutes of recorded lecture transcribed during one minute of participants working time. This is a quantitative parameter of output. In the instructions we explicitly required to "transcribe as much as possible", which is though not specific, but still represents a maximization goal.

This is the main goal of the work, while quality parameters can be seen as the restrictive boundaries altering the amount of recording one can transcribe. While we do not specify in the instructions whether the participant is the only person responsible for the completion of the transcript, we can suggest that if she sees herself is the only responsible, then the importance of quantity will definitely rise. If the participants believe that she is doing only part of work, then the quantity is less important.

Speed of Typing

Speed of typing is the final parameter which we want to discuss. Speed of typing measured as the number of words typed by the participant in one minute of her working time. This is a quantity output parameter not directly prescribed by the instructions. However, the amount of words typed is an important parameter of observable quantity. It is strongly related to the typing abilities of each participant, but also can be influenced by other quality and quantity output parameters.

In general this parameter can be seen as the one reflecting the overall effort of the participant. In fact higher speed of transcribing is associated with higher speed of typing (without making any directional link here). Also increase in preciseness, if it does not result in significant loss of speed of transcribing, can mean to the higher amount of words typed in a minute.

Formatting

We also measure the formatting activities of the participants: how many changes have been introduced to the document upon the receipt of the instructions. In particular we count the number of units which were "inserted", "deleted" and "formatted" as provided by the MS Word function "Compare texts".

4.2.3. Relationship between main output variables

We should keep in mind that the transcription process consists of listening to the recording, pausing the recording, typing the words one has heard, moving recording backwards to the point where the typed piece starts, playing recording and so on. As a result, increasing preciseness of the transcription means higher frequency of stops and move-backward-move-forward activities. It definitely means less recording is heard through within a certain period of time (e.g. one minute), but not necessarily less words typed (for a "looser" transcript one listens more and writes less). As for change in the accuracy of transcript it signifies the loss in both typing and transcribing speed.

At a more general level we can see the speed of typing as a quality output variable, change in which represents pure effort (also learning effect) of the participants (if other parameters are held constant).

Significant change in preciseness should be negatively related to the speed of transcribing, but it can both increase and decrease speed of typing. Significant change in the accuracy of typing can negatively influence both speed of transcribing and speed of typing. Significant change in speed of transcribing can be achieved only by decrease of preciseness and / or accuracy. Thus, significant increase in the speed of typing can be a result of deterioration of accuracy and / or increase in speed of transcribing without significant change of preciseness.

Here we want to stress that in the beginning of the experimental session, during the control time, the only available information for the participants was about the necessity to provide a transcript of the lecture. At that point the participants could form some beliefs about what it meant to make a transcript, including the beliefs about how precise it should be, what level of misspellings and typos can be tolerated and how important those parameters in comparison with quantity are.

We believe that the focus cannot be put on a large number of issues simultaneously, that is why the difference in the instructions provided at the main time could be able to limit the scope of options and move the focus towards a single parameter of output, different for each treatment group.

For example, in the first group there is no information distracting participants from the requirements set in the instructions. That means first of all the requirement for maximisation of the preciseness. Having this achieved the overall goal is set as maximising the quantity of the transcription. In such situation those participants who already had the quality requirement fulfilled because of their quality standard, can use effort to increase the speed of transcribing. Others though can re-evaluate the

situational importance of the preciseness and increase it. Increase in the importance of preciseness will generally mean the decrease in the importance of speed. That is why even remembering the necessity to provide as much as possible, sufficiently large increase in preciseness can also lead to some decrease in speed. General information about the purpose of the task received by the participants in this group also should lead to some improvement in the accuracy, but this can come only for those who can "afford" it without losing too much speed.

In the group with identified customer the first focus should be not only on preciseness, but also on the speed. If this double focus intensifies the trade-off between preciseness and speed, then extra effort can be invested in speed.

In other words if we see the ideal subjective standard for each parameter as stable within short time span then receiving the treatment information may change the relative situational importance of quality versus quantity for participants. Significant increase in the importance of one of the parameters can be achieved only by significant reduction of the importance of the other. In the absence of additional effort (due to the increased motivation[33]) this should inevitably lead to the improvement in one parameter at the expense of the other. However, in the presence of the increase in effort it can be channelled towards improvement of target parameter without deteriorating other parameters.

That is why in the context of the selected real-effort task we hypothesize the following:

1. The explicit requirement for qualitative parameter (preciseness), stated in the instructions, will increase the situational relative importance of this parameter for all participants.
2. Requiring the disclosure of the employee's name will result in the increase in the relative situational importance of visible qualitative parameters (accuracy of transcript, formatting).
    a. Social pressure leads to the increase of the effort (real or simulated) which should be clearly recognizable. Simultaneous increase in preciseness (required quality) and accuracy (visible quality) can lead to the decrease of quantity parameters.
3. Providing the information about the customer will lead to the increase in the importance of the quantity of output (speed of transcribing or/and speed of typing)
    a. Empathic concern leads to the increase of the effort which should be valuable from the point of view of the customer.
4. Providing both the information about customer and requiring the provision of employee's name will lead to the simultaneous improvement of accuracy and speed, which can result in insignificant improvement of each of the parameters.

---

[33] Also due to learning effect, but one can argue that learning itself consumes effort and thus requires motivation.

4.3. Questionnaire-based measures used in the experiment

4.3.1. General overview

Due to the use of several pre- and post-questionnaires the quantitative and qualitative results of the experiment can be related to a number of descriptive variables. Such variables can be grouped in the following way:

1. Parameters characterising the level of either dispositional or situational empathy
    a. Empathic trait – parameter of dispositional empathy;
    b. Empathic state – parameter of situational empathy;
    c. Distress state – parameter of situational distress;
2. Parameters characterising the beliefs about helping behavior of others directed towards the people with disabilities (serve as approximation for agreement with prescriptive and descriptive social norms of helping)
    a. Prescriptive norm – level of agreement with the necessity to help people with disabilities;
    b. Descriptive norm – level of agreement with the fact that others help people with disabilities;
3. Parameters of task proficiency and interest in the task
    a. German knowledge – level of German proficiency
    b. Task easiness – level of agreement with the statement that the task was easy
    c. Task interest – level of agreement with the statement that the task was interesting

In order to reflect both approaches to understanding the empathy – both as persistent individual trait and as context dependent emotion – we have introduced the participants to two different questionnaires.

4.3.2. Parameters characterising the level of dispositional and situational empathy

Most current approaches admit that empathy has both affective and cognitive components and it is important to distinguish between them (Edele et al, 2013). Cognitive aspect relates to understanding of the others' mental states (perspective-taking, mentalizing), while affective empathy relates to feeling with other or experiencing feelings congruent with other's emotions.

In their study affective empathy (state-like) is the strongest predictor of altruistic sharing. Not the proper understanding of other's perspectives (cognitive empathy) but the disposition to react emotionally towards their plights seems to account for altruistic behaviour in dictator game.

Batson et al (2007) examine the role of valuing the welfare of the person in need as an antecedent of empathic concern. They design a new model of antecedents of empathic concern with perceiving the

other as in need and valuing her welfare as two separated sources of empathic concern. Adopting other's perspective lies on the path from valuing to empathy. In general, perspective taking can lead to empathy without valuing welfare; however the researchers argue that in the absence of direct instructions to adopt other's perspective, valuing can influence empathic concern without perspective taking, while when the other's welfare is valued, perspective taking is spontaneous.

The valuing can be increased by means of reducing prejudice, improving attitudes towards out-groups or introducing more cooperation. This in turn can lead to prosocial and even altruistic motivation.

Gehlbach et al (2012) working with the concept of social perspective taking (SPT) come up with a list of at least thirteen specific factors that impact motivation to participate in SPT. These factors included both those which motivated SPT, and those which inhibited it. Among the first was for example the consideration about the importance for SPT (stakes for the other), intrinsic interest in SPT or hope to acquire self-knowledge. Interestingly such factor as seeing SPT as part of the role or identity (necessary for some occupations) could have mixed effect on SPT. The same is true for familiarity: sometimes people are motivated to take the perspective of people from different cultures. At the same time a lack of energy and being under cognitive load serve as factors inhibiting SPT (which for can support the idea that it is a mentalizing and not empathizing / emotional activity).

Indeed according to Schulz et al (2014), decisions are driven by two different modes of cognitive processes: one fast, automatic, effortless and often emotionally charged and the other deliberate, requiring greater cognitive capacity (e.g. Kahneman, 2013). So the researchers ask whether other-regarding behaviour is rooted in the affective system or it is rather an effortful cognitive process.

They conducted an experiment with two parallel tasks – one social (mini dictator games) and one cognitive load (remembering the sequence of letters and pressing some button every time they here the same letter that resounded two letters before).

They found that subjects in high-load condition are more generous on average (chose the fair allocation more than 43 percent of time – out of 20 mini DGs – compared to about 31 percent the low-load condition). Thus, affective system plays an important role in altruistic choices. They conclude that affective system is associated with heuristic and links decisions towards altruistic choices. Basic altruism is interpreted as a fast decision heuristic. Inequity aversion seems to require more cognitive resources than simple generous behaviour. Thus, the deliberation system adjusts behaviour in self-regarding manner, at the same time moderating affective reaction so that it would be more tailored to the particular situation.

While keeping in mind the complexity of the construct of empathy, we needed an instrument able to measure it in our experimental settings without priming the participants or making this a focal point of their effort. If we remember that questionnaires were not the primary, but rather secondary instrument

in our research design it was important to reduce the time participants spent for answering them. That is why we were looking for the shorter versions of normally rather long psychological scales.

In particular for measuring dispositional empathy we used the Empathy Quotient (Baron-Cohen et al, 2003) in its shortened version of eight-item form of the Empathy Quotient (Loewen et al, 2009). Originally Empathy Quotient was developed within the theory of gender differences positing that men are more adept than women at systemizing and women are better at empathizing. Empathizing is defined here as the ability to identify and respond to other's emotions. The researchers designed the empathy quotient with the purpose to combine affective, cognitive and mixed components in one measure of empathy. Affective component is about feeling an appropriate emotion triggered by another's emotion; cognitive component includes understanding and/or predicting what someone else might think, feel, or do.

While original scale of Baron-Cohen et al (2003) consists of 40 items, in 2006 the group of researchers including Baron-Cohen attempted to reduce it to 22 items (Wakabayashi et al, 2006). They chose the items with loading of above 0.40 and achieved internal consistency (Cronbach's alpha) of 0.90 for these high loaded items.

Further Loewen et al (2008) argue for further reducing the questionnaire in order to better match the constraints and research approaches of social sciences other than psychology.

What is even more important for us is that the researchers tested their reduced form of the Empathy Quotient in relation to the empathic capacity predicting an individual's propensity to give money to the charity. They used a large sample and more general population than was used in previous studies (Loewen et al, 2008). For the eight item version of the survey four affirmative EQ questions with the highest principal component factor loadings and the four reversal items with the highest factor loadings were chosen. Internal reliability of this scale was found to be reliable at Cronbach's alpha =0.76. Both EQ-Short and EQ-8 supported the differences between man and women in terms of empathizing scores, with women being significantly more empathizing then men.

Researchers also found in their internet survey that people with higher empathy scores reported giving more money to the charity (however, there is no data on whether the respondents actually gave money to the charity).

In order to assure consistency between questionnaires provided within the experiment we used the Likert's scale asking participants to indicate on a 7-point scale (from not at all to very much), how well each of eight statements describes them. Importantly we mixed the positively and negatively framed statements.[34]

---

34 In calculations we tested to ways of aggregating data: first, using averages of the scores received on Likert scale, and second, adjusting the measure by means of transforming the answers of 6 to 7 into scores equalling 2 for strong agreement,

As it has already been mentioned in the literature review, one of the most prominent researchers working in the field of empathy is Daniel Batson, who already in the 1970s came up with the idea of empathic concern and conducted multiple experiments testing the link between empathic concern and helping behavior.

In the paper Batson wrote with his colleagues in 1978 (Coke et al, 1978) he proposed a two stage model of empathic meditation of helping. In this model he tries to combine two perspectives, one seeing empathy as emotional and another as cognitive process. The cognitive perspective argued that by taking the perspective of the other a person can see the world through the other's eyes, which triggers helping behavior. Emotional view of empathy was related to the psychological arousal and emotional response to the sufferings of the person in need, which motivated helping.

For Batson and colleagues (Coke et al, 1978) the cognitive aspect was a driver for emotional response, which in its turn was supposed to lead to the helping behavior. In other words first people take the perspective of the other, which ads to empathic emotional response, which increases motivation to see the person's need reduces.

In order to test the model Batson and colleagues conducted multiple experiments. Already in the early papers (Coke et al, 1978) we can find a measure of empathic concern developed by Batson together with his colleagues in 1976.[35] Since the understanding of empathic emotion led to some controversy about its roots, it became very important to distinguish between other and self-directed\response to the other's plight. In particular one can experience empathic concern, which directs the behavior towards reduction of the distress of the other person in need and on the other hand one can experiment a feeling of personal distress, which directs the behavior towards reducing own distress and can result in helping it is the most effective mean for reducing own distress. Hence, in their measurement instrument the researchers distinguish between empathic concern and personal distress. In the question nary used for calculating the Index of empathic concern and personal distress the researchers include the number of adjectives which they believe are relevant to the emotional states of empathic concern and distress and ask their subject to rate their level of agreement with them at specific point in time (usually after some experimental manipulation directed at producing empathic response of participants). The first list of adjectives included as possible measures of empathic concern such words as moved, soft-hearted, sorrowed, touched, empathic, warm, concerned and compassionate. For measuring personal distress such adjectives as alarmed, perturbed, disconcerted, bothered, irritated, disturbed, worried, uneasy, distressed, troubled, upset, anxious and grieved were initially used. As a result of further studies the list of adjectives was continuously refined. In particular in the Coke et al (1978) five adjectives measuring empathic concern (soft-hearted, empathic, warm, concerned and

---

answers of 3 to 5 into scores equalling 1 for weak agreement and answers equal or below 2 into scores of 0 for disagreement. Of course we took into account the positive versus. negative phrasing of the statements.

[35] Here the following reference is used: Batson, S.D., McDavis, K., Felix, R., Goering, B., & Goldman, R. Effects of false feedback of arousal on perceived emotional state and helping. Unpublished manuscript, University of Kansas, 1976

compassionate) and three adjectives qualifying for the index of personal distress (upset, alarmed and troubled) were found to have the highest loadings in the factor analysis.

In the other studies (Toi and Batson, 1982; Batson et al, 1983,) researchers drop such adjectives as empathic and concerned and reduce the list to eight adjectives measuring the empathic concern and six reflecting feeling of empathy. They show that the answers load highly on separate orthogonal factors. Factor analysis produced a clear two-factor solution accounting for 67 percent of the variance and all eigenvalues above 1.0.

The resulting emotional response scale (Batson et al, 2007) includes six emotions assessing the feeling of empathic concern: sympathetic, soft-hearted, warm, compassionate, tender and moved. Distress is measured using eight adjectives: alarmed, grieved, upset, worried, disturbed, distressed, troubled, perturbed. Previous experiments showed high internal consistency of the scale with Cronbach's alpha =0.94 for distress index[36] (Batson et al, 1983) and Cronbach's alpha =0.90 for the index of empathic concern (Batson et al, 2007).

We used this set of 14 adjectives to measure emotional response to the situation our participants find themselves in. We offer them to rate on the 7 point Likert scale their level of agreement with the statement that at the present moment they experience each particular emotion. We mixed the adjectives in the list in order to reduce the chances of guessing for participants.[37]

We introduced the measure of social value orientation in our analysis introduced by Murphy and colleagues (Murphy et al, 2011) and since then actively used in economic and behavioral research. We used a SVO slider in form of a paper based choice task with six primary Slider Measure items. Each item represented a resource allocation choice over a well-defined continuum of joint payoffs and a decision maker has to indicate her hypothetical allocation choice on the line. The rationale and mechanism being SVO slider is described in Murphy et al (2011).

---

[36] Using six adjectives (without distressed and troubled)
[37] As advanced knowledge of English was a pre-requirement for the participation in our experimental study we decided to use the original adjectives in English. In the future studies, though, the German translations can be also provided.

V. Discussion of the Results

5.1. Overview of structure

We start the chapter from general analysis of the data, providing descriptive statistics and tests of normality. We do this for the main output parameters in the control time and in the main time. This allows us to determine the main statistical methods which are going to be further used for statistical analysis.

At the next step we move to the analysis itself. Based on the research questions and hypotheses set in the thesis the major tasks of the statistical analysis were to answer the following blocks of questions:

1. What kind of trade-offs face the employees by performing a multiattribute task in our experimental settings? What are these trade-offs in the control time (before employees get informed about the purpose of the task and its beneficiary)? In the main time (after the provision of the above information)? Do the trade-offs in the main time differ across the treatment groups?

2. Do the employees change their trade-off upon receiving treatment information? In other words, do employees significantly change the focus towards specific attributes of the task in the main time compared to the control time? If yes, which attributes do get in focus? In which groups?

3. Do individual characteristics of the employees interrelate with the behavior in our experimental setting? In particular do the empathic ability and the level of the agreement with the specific norms move the trade-off points within the treatment groups? Does the task proficiency play a decisive role by setting the trade-off points? For all treatments?

4. Does the structure of final output significantly differ across the groups, taken into account the individual differences among participants? How much can the differences in the output be attributed to the treatment effects?

The first set of questions and the related analysis is of the explorative nature. At this stage our goal is to throw a first look at the relationship between different attributes of task and their importance to the employee, without making strong arguments regarding causality. Answering these questions allows us to draw conclusions and make relevant interpretations of the next steps of the analysis. We use correlation analysis as the statistical method for this part.

As a result we observe a strong difference in opinions regarding the relative importance of preciseness of the transcript and speed of transcribing observed in the control time. In the main time it is, however, characteristic only of the behavior in Group 2 (group with anonymous employees having information about the customer) where the negative relationship between these two parameters of output is significant ($r=-0,43$, $p<0.05$). This means that within this group employees differ in their evaluation of the importance of preciseness versus speed of transcribing. In the two groups with open employee

identity preciseness is strongly positively related to the speed of typing, thus linking it to the task proficiency.[38] Moreover in the group with combined treatment effects more proficient participants (higher speed of typing) were also the ones having better transcribing speed and fewer mistakes.

Thus, we can conclude that proficiency played a significant role in achieving results in the groups we open employee identity.

At the same time the initial trade-offs played a significant role in determining the trade-offs in the main time. In general the level of output parameters reached in the control time strongly correlated with the level of the same parameters in the main time. [39]

The second set of questions allows us to see whether the treatment interventions were able to change the intimal trade-off points of the employees. Due to the differences in the experimental instructions we expect that the new trade-off points do significantly differ from the initial ones (achieved in the control time), and also differ across the groups. We use the related samples analysis (both non-parametric and parametric) for this part.

As a result we see that at the aggregated level provision of the task purpose as well as other treatment information led to the increased amount of effort put into significant improving the parameters of quality of output (with $p < 0.05$ for both parameters). Importantly this did not significantly influence the quality of output on aggregate level. At the level of individual groups we observe that the aggregated results regarding preciseness are driven by Group 1 and for the amount of the mistakes – by Group 3 (and to a much lesser extent Group 4).

Combining the results of these first two phases of analysis we can say that general information about the purpose of the job and the general customer led to the focus on preciseness; adding the information about the selected customer led to a divide between the focus on preciseness and speed of transcribing; opening the employee's identity moved the focus from preciseness to the reduction of mistakes (improvement of preciseness is seen as secondary and varies with the typing skills); and finally combing all treatment effects led to the balanced focus on different output attributes with slight preference towards reducing the mistakes and with strong interrelation between task proficiency and output.

The third block of questions brings us to a more complex picture of reasons for the behaviour within groups. We know that the related samples analysis which we conducted for answering the second set of questions allowed us to compare the behaviour of participants between control and main time while taking account of idiosyncratic differences among participants. At this point we want to put them back into analysis and explore how these individual differences in preferences and abilities (task

---

[38] Importantly at this point we confirm our interpretation of the parameter of speed of typing as the one reflecting the task proficiency of the employee.
[39] As exception can be seen only the parameter of preciseness in Groups 1 and 2.

proficiency) relate to the trade-off points for the individual output attributes for each treatment group. In particular in Group 2 we discover that an agreement with the descriptive norm of helping significantly positively related with preciseness (r=0,51, p<0,05) and significantly negatively related to the amount transcribed (r=-0,42, p<0,05). This explains the negative relationship between these parameters which we observed in the beginning in the correlation analysis. Now we can make a conclusion that in Group 2 the participants split in their opinions regarding the importance of preciseness versus speed of transcribing and do it on the basis of their believes about the helping behaviour of others. We interpret it in the following way: if the employees believe that others do not help, then they feel themselves responsible for providing the customer with as much output as possible. Based on our conceptualisation of different parameters of output we suggest that if participants put themselves into the customer's shoes they realise that amount of the recording transcribed is the output parameter which customer himself cannot control and that is why it is very important. That is why perceived single responsibility of the output leads in this group to the larger amount of the recording transcribed with lower preciseness. If the responsibility is seen as shared then participants pay more attention to the preciseness and lose in the speed of transcribing. Interestingly more empathic participants in this group make more precise transcripts, without losing speed of transcribing.

In the fourth part we combine the results of previous research to get a complex picture of the behaviour.

By that point we already have an understanding of the relationship between attributes; we know whether and how the behavior within groups changed due to the treatment effects; we have information regarding the influence of the individual parameters within the groups. However, what we what to see – and what is crucially important for testing our hypotheses - is whether the final output of different treatments is indeed significantly different. We need to look for the changes not only within the treatment groups, but also among them. We assume that within each parameter of output we can compare its average levels for each of the treatment groups and the resulting differences can be attributed to the treatment effects. At the same time such comparison cannot be made without taking account of the individual differences in the beliefs regarding the trade-offs and individual differences in the task proficiency. As we have shown above there is a strong relationship between the level of output for each of the attributes in the control and main time. And the initial level undoubtedly has an individual component in it. This component combines the beliefs about the relative importance of different attributes of output and proficiency of the participants. Moreover in the previous analysis we discovered that such parameters as the empathic trait and adherence to descriptive norm play a significant role in selected groups. That is why we consider it important to take account of these variables in the comparison analysis. To do that we used the regression analysis.

Conducting the regression analysis gives us the opportunity to analyse the output across groups taking account of the initial differences in the trade-offs and individual characteristics, thus narrowing down the range of possible reasons to the treatment effects.

As a results of this analysis we come to the following conclusions

- Group 1 not only was the only one with exclusive focus on preciseness, but this was also the group, which achieved the highest level of preciseness among all groups (with significant difference to Group 3); suffered from high amount of mistakes (significantly higher than in Group 3 - leader in accuracy), and low amount transcribed (significantly lower than Group 2 leader in the amount transcribed)[40].

- Group 2 strongly varied in the attitude towards preciseness, but at the end the preciseness of their transcripts was not significantly worse than those of Group 1 (leader in preciseness), while the chances to transcribe more recording were significantly higher in this group than in Group 1. The amount of mistakes was significantly higher than in Group 3 (leader in accuracy).

- Group 3 not only was the one with most significant focus on reduction of mistakes, but in the end it also achieved the lowest level of mistakes among all groups (with significant difference to Groups 1 and 2). This however led to the loss in the amount typed, which is the lowest among all groups (and significantly lower than in Groups 1 and 2).

- Group 4 strongly varied in the level of output they reached depending on the task proficiency of the participants. However, none of the attributes had the level significantly worse than the leader group for the respective parameter. We confirm that the focus of this group is rather balanced.

5.2. Demographics of the experimental sample

100 students of the Faculty of business, economics and statistics of the University of Vienna took part in the experimental sessions. Several students submitted not fully filled questionnaires, which is why for each of the four main output variables there are 100 observations, but for several descriptive variables number of valid observations varies from 93 to 97. The Table 2 gives an overview of the values for main demographic variables. In particular we can see that the age of participants ranged from 18 to 41 years with mean value around 23 years. In fact almost 80 percent of participants were 25 years old or younger.

Interestingly most frequently participants were in their second semester (24 percent), fourth semester (15 percent) or sixth semester (19 percent), which means that most of them were undergraduate students. About 60 percent of participants had some previous working experience with slightly more than 20 percent having more than 2 years of experience. On average the participants were in their fifth

---

[40] Here and further we either use the results that are taken from the parameter estimates of the respective models and based on the values of the B-coefficients and their significance; or the estimates of the means for the respective parameters and their pairwise comparisons and related significances.

semester of studies at the University of Vienna and they had about a year of work experience. We can also add that 56 percent of participants were female and 44 male, which means that gender split was relatively balanced.

Descriptive statistic for demographic variables

		Age	Semester	Work experience	German knowledge	English knowledge	Task easiness	Task interest
N	Valid	93	93	93	95	97	97	93
	Missing	7	7	7	5	3	3	7
Mean		23,28	5,06	1,80	3,46	2,89	3,92	4,02
Median		23,00	4,00	1,00	4,00	3,00	4,00	4,00
Std. Deviation		3,63	3,22	2,86	,71	,43	1,70	1,83
Minimum		18	1	0	2	2	1	1
Maximum		41	18	17	4	4	7	7

The requirements set in the announcement included excellent knowledge of German and very good knowledge of English. As a result the average level of German knowledge reported by participants were equal to 3,5 (with "3" being an advanced speaker and "4" – a native speaker) and the average level of English knowledge reached 2,9 (with "2" being an intermediate speaker and "3" – an advanced speaker). In general almost 60 percent of participants had German as their mother tongue and more than 80 percent considered themselves advanced speakers of English (another 4 percent had English as their mother tongue). These results provide support for the statement that participants were proficient enough to perform the task: they had sufficient knowledge of English to understand the instructions and of German to understand the recorded lecture and be able to transcribe it.

Participants' agreement with the statements regarding the easiness of the task and whether the task was interesting were evaluated on the Likert scale using scores from 1 (not at all) to 7 (very much). In fact about 22 percent of participants considered the task very difficult (scored 1 or 2 for the agreement with easiness of the task) and another 20 percent considered it difficult (scored 3) while 23 percent though that task was very easy (scored 6 or 7).

As for the intrinsic qualities of the task, about 25 percent of participants found the task not interesting at all (scored 1 or 2 for the agreement with the statement that the task is interesting), while the same number of participants – 25 percent – found the task very interesting.

While in both cases participants used the whole range of answers, the average agreement with the statement that the task was easy was at the level of 3,9 points and with the statement that the task was

interesting at the level of 4 points. Medians for both parameters were equal to 4 points. This means that generally the participants considered the task rather easy (but not very easy) and rather interesting (but not very interesting).

5.3. Descriptive statistics for the main output parameters and tests of normality

5.3.1. Analysis of the parameters in the control time

We start from the analysis of the initial output variables. As we see from the Table 3 on average in the first 10 minutes participants typed 191 words and transcribed 3 minutes of recorded lecture, they used 80 words to transcribe one minute of recorded speech and made about 9 mistakes per 100 words of typed text.

Table 3

Descriptive statistics for output parameters in the control time[4142]

	Preciseness_10	Mistakes_10	Speed_Minutes_10	Speed_Words_10
Mean	80,436	9,333	,292	19,112
Median	85,870	5,724	,215	17,500
Std. Deviation	27,284	8,243	,221	8,594
Variance	744,421	67,951	,049	73,861
Skewness	-,644	,998	2,236	1,138
Std. Error of Skewness	,241	,241	,241	,241
Kurtosis	,257	,000	6,305	2,333
Std. Error of Kurtosis	,478	,478	,478	,478

We should be very careful in describing these parameters of output in terms of means, as figures for skewness and kurtosis suggest a pattern of non-normality of the data. Except for the measure of preciseness of the transcript, all other output variables are characterised by positive skewness, which means that the data has a tendency to cluster to the left – heavier tail in the area of low numbers. The

---

41 Here and further for distinguishing between the parameters of output reached in the control time and in the main time, the former are given in the tables with extension "10" (as control time last 10 minutes), and the latter without any extension.
42 Here and further in the statistical tables:
- Variable "Preciseness" refers to the parameter of Preciseness of the transcript measured as the number of typed words used for transcribing one minute of the recording;
- Variable "Mistakes" refers to the parameter Number of Mistakes measured per 100 words of types text;
- Variable "Speed_Minutes" refers to the parameter Speed of Transcribing, measured in the minutes of the recording transcribed within one minute of working time;
- Variable "Speed_Words" refers to the parameter Speed of Typing measured as the amount of words typed within one minute of working time.

parameter of preciseness also has a significant skew, but to the right side of distribution. Variables characterising the output quantity are also characterized by the significant kurtosis.

Normal Q-Q plots for output parameters in the control time
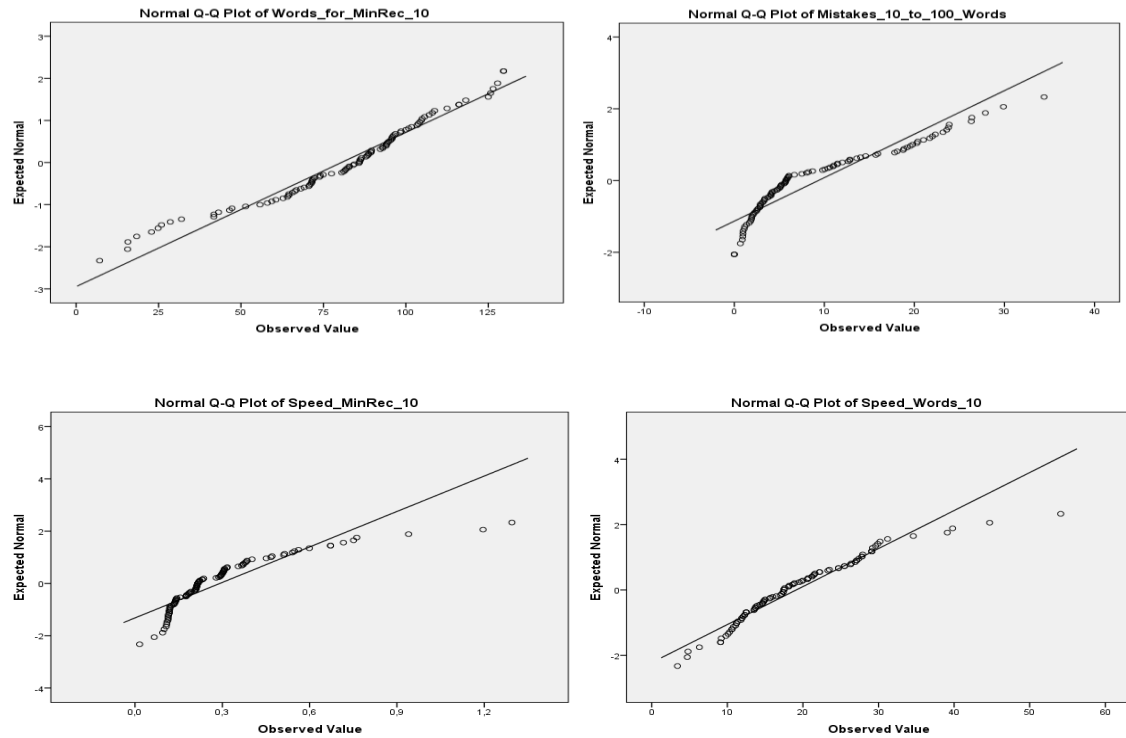
Since the normality of data has important implications for all further analysis, we also check normality visually by means of Q-Q plots (Figure 5) and further with Kolmogorov-Smirnov and Shapiro-Wilk tests (Table 4).

Test of normality for output parameters in the control time

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	Df	Sig.	Statistic	Df	Sig.
Preciseness_10	,102	100	,012	,958	100	,003
Mistakes_10	,214	100	,000	,872	100	,000
Speed_Minutes_10	,195	100	,000	,777	100	,000
Speed_Words_10	,102	100	,012	,933	100	,000

a. Lilliefors Significance Correction

Q-Q plots show the deviation from the normality for all variables, which is supported by the values of both Kolmogorov-Smirnov and Shapiro-Wilk tests (Table 4), all of them being significant at p<0,05 level.

5.3.2. Analysis of the parameters in the main time

Moving to the analysis of output variables in the main working time after the treatment manipulation has been introduced, we first analyse them without division into groups, but using the whole sample of 100 people. We start from the analysis of the initial output variables.

Starting with overview of averages for the output variables in the main time we can also compare them to the averages for the control time. As in the case of the control time, the main time variables are also non-normally distributed (figures in the Appendix 8 show significant values of the Kolmogorov-Smirnov and Shapiro-Wilk tests for most of the parameters). On average in the main time participants typed 188 words and transcribed 2,6 minutes of recording in every 10 minutes of their work (Table 5). They typed 85 words per minute of recorded lecture and made 7,5 mistakes in every 100 typed words. If we compare these results with those from the control time we can observe a slight improvement in parameters of output quality: increase of average preciseness from 80 to 85 words per minute of recording and decrease in the amount of mistakes from 9,3 to 7,5 in 100 typed words. Average quantity of output (either in minutes of the recording transcribed or in number of words typed in one minute of the working time), though, almost did not change.

Table 5

Descriptive statistics for output parameters in the main time

	Preciseness	Mistakes	Speed_Minutes	Speed_Words
Mean	84,680	7,490	,2578	18,772
Std. Error	2,263	,670	,018	,708
Median	94,225	4,831	,216	18,540
Std. Deviation	22,630	6,696	,181	7,081
Maximum	10,58	,000	,055	4,30
Minimum	105,98	27,968	1,140	38,40
Range	95,40	27,968	1,085	34,00

This trend is also obvious if we compare the medians (instead of means) of the parameters in the main time to the ones in the control time: 185 versus 175 words typed per 10 minutes; 94 versus 86 words

used to transcribe a minute of recording; and 4,8 versus 5,7 mistakes per 100 typed words in the main time compared to in the control time.

What is also interesting and important for the further analysis is the range of the parameters. While preciseness on average (or using median) was in the range between 85-95 words for 1 minute of lecture, some participants were typing only 10 words per 1 minute of recording and some typed more than 105 words. This difference means that while participants with high level of preciseness were typing each and every word they could hear in the recording, at the other extreme participants with low preciseness were typing only a small proportion of words they could hear, missing some parts of the sentences or even the whole sentences (or in some cases even paragraphs). Those with very low preciseness are much further from the central tendency parameters than those with high preciseness. This means that they are more of "outliers".

The similar situation can be observed for the speed of transcribing where some participants were transcribing only 30 seconds of recording in 10 minutes, while others 11,4 minutes. In fact in order to "cover" the amount of recording longer than the amount of time provided one obviously should be listening only to a part of the recording and rewind other parts of it. Since larger values are located much further from central tendency parameters than the small values, we can suggest that there are several extreme observations including participants with very high speed of transcription and low preciseness.

Taking these two observations together we can suggest that there are several individuals in the sample who provide very high speed of transcribing, but they achieve it due to the tremendous loss of preciseness of the transcript.

Also the accuracy of the transcript differs dramatically across participants, with central tendency parameters lying closer to the low end of the range. While some participants made 27 mistakes per 100 words typed and others did not do any, on average participants made 7 mistakes (median lays by 4,8 mistakes) per 100 words of transcript.

Interestingly speed of typing is the only parameter where mean and median are very close and lie in the middle of the range. Analysis of the distribution of output variables in the main time by means of the box-plots (Figure 6) supports our conclusion that the participants who were extremely high on speed of transcribing were mostly the ones who made quite loose transcripts (low preciseness)[43].

However, they were not the same participants who lacked accuracy of typing (made a lot of mistakes). That means that some participants were intentionally trying to provide as much transcript as possible sacrificing the preciseness for that, but these people were not necessarily the ones having most of the mistakes. This once again shows that preciseness and accuracy are two different dimensions of quality perceived differently by the participants.

---

[43] They have same identification numbers for outliers in the box-plots.

Figure 6

Box-plots for output parameters in the main time
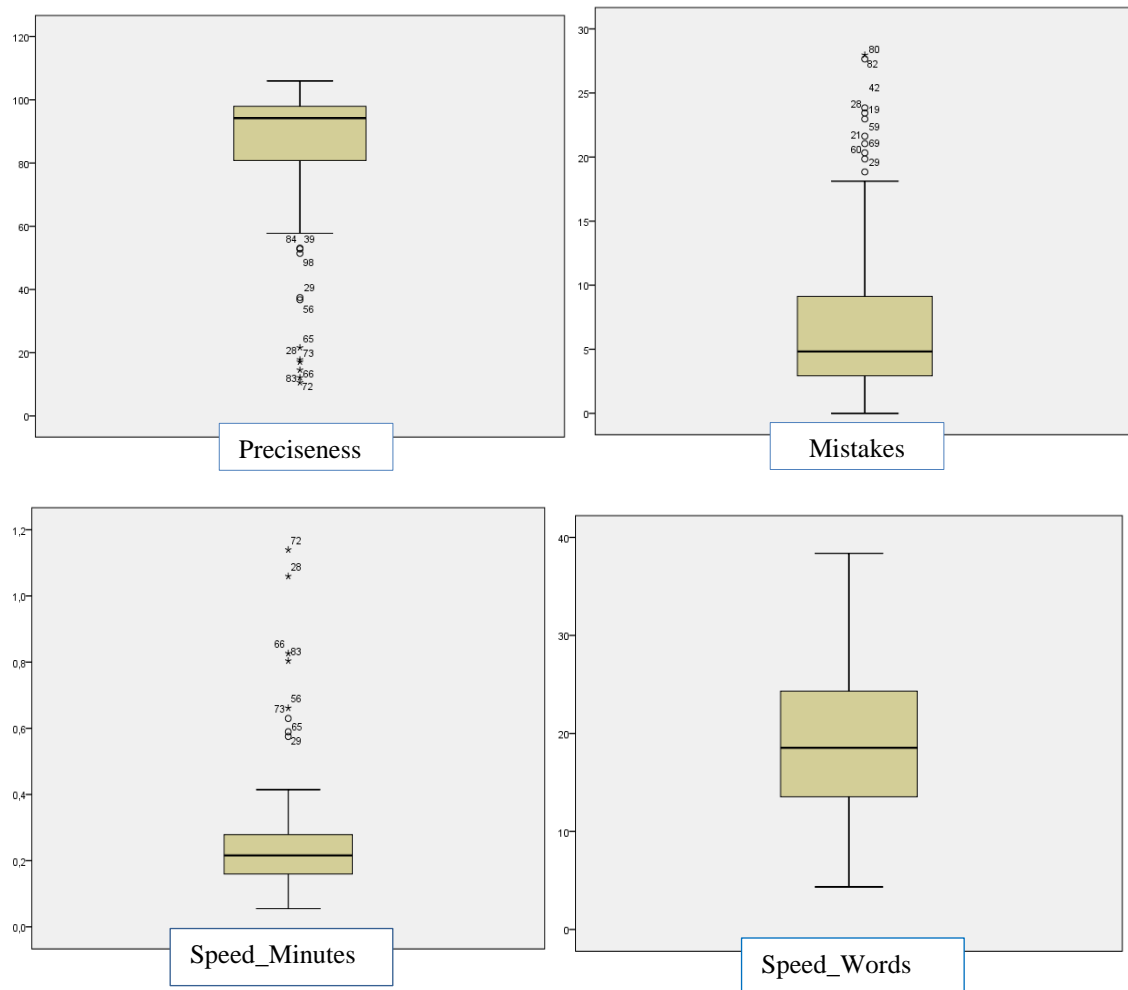
By the analysis of the extreme values we should remember that these are not outliers in their ordinary sense. These values represent the output which is the result of 50 minutes long effort and cannot be a result of pure chance. They reflect either the preferences or beliefs regarding the relative situational importance of the output variables.

With relatively modest sample outliers can mean that there is a stable number of such cases in the population. In particular we have around 10 outliers for each output parameter except the speed of typing. In a general sample of 100 people it constitutes about 10 percent.

This brings us to the conclusion that the specifics of the data require particular robust methods which could account of the non-normality[44], in particular:

- We need to use non-parametric analysis which is not based on the parameters of the mean values

---

[44] One way to correct the situation would be either delete the cases with outliers or to change the figures for outliers for the ones within two standard deviation from the mean. However these actions are all built on the premises that these cases are also outliers in the population, which, we believe, is not necessarily the case (as explained above).

and, thus, less influenced by what happens in the tails of the distribution;

- In cases where we use parametric analysis we compliment it by using the method of bootstrapping.

Summing up the results of the general analysis of output variables in the main time we stress are the following issues:

- Data for output variables in the control time are non-normally distributed.
- Data for output variables in the main time are non-normally distributed.
- The distributions for output variables in the main time contain extreme values which comprise about 10 percent of the sample for each output variable.
- Due to the specific nature of the data and comparatively large number of extreme values robust methods are to be employed in the analysis, in particular non-parametric methods are to be complimented by parametric analysis, which results are corrected using bootstrapping technique as the way of assuring higher robustness of the results.

5.4. Relationship between output parameters in the control and in the main time

5.4.1. Analysis of the parameters for the whole sample

From the description of the output variables in the previous part it is clear that they are also highly interrelated. That is why having checked the properties of the data, we can quantify the relationships between the main output variables (Table 6).[45]

Table 6

Correlations between output parameters in the control time

	Speed_Minutes_10	Preciseness_10	Mistakes_10
Speed_Words_10	**,697**[**]	-,041	-,128
Speed_Minutes_10		**-,632**[**]	,083
Preciseness_10			**-,231**[*]

    *Correlation is significant at the 0.05 level (2-tailed).

    **Correlation is significant at the 0.01 level (2-tailed).

First, we observe a significant positive correlation between two parameters of output quantity– speed of typing (in words typed) and speed of transcribing (in minutes of recording transcribed) with r=0,697 (p<0,01). Second, preciseness of transcript – parameter of output quality – is significantly negatively correlated with speed of transcribing with r= -0,632 (p<0,01), but virtually uncorrelated with speed of

---

[45] Here and further taking into account the non-normality of the data we always analyse the relationship between output variables using non-parametric correlation parameter Spearman's Rho.

typing with r= -0,041 (p>0,05). Third, amount of mistakes is uncorrelated with any of the variables measuring output quantity and slightly negatively correlated with preciseness of the transcript (r= -0,231, p<0,05).

Interpreting these correlations we should remember that there are several underlying dimensions which can influence each of these parameters, in particular individual proficiency of each participant in the task, including typing skills and language proficiency. Although typing abilities have not been measured directly, speed of typing is obviously restricted by this parameter. Thus, if higher speed of typing is associated with higher task proficiency, then the significant correlation between two quantitative parameters of output is quite logical.

Additionally beliefs of the participants about relative importance of preciseness and quantity maximization influence the trade-off between these parameters of output. The negative correlation between preciseness and speed of transcribing means that those employees who transcribed more recording made it less precise and that is why those who paid more attention to the requirement of preciseness ended up with less recording transcribed compared to their "faster" colleagues.

Finally negative correlation between preciseness and amount of mistakes means that higher preciseness can be associated with lower amount of mistakes, which generally can mean that if one values the importance of preciseness, she is more likely to make a more accurate transcript in terms of misspellings.

Table 7 renders some support to our understanding of the relationship between output parameters by means of exploring their correlations with selected individual characteristics (questionnaire-based), namely German proficiency and perceived task easiness. In particular we see that speed of typing significantly positively correlates both with language knowledge (r=-0,398, p<0,01) and with self-reported measure of easiness of the task (r=-0,328, p<0,01), meaning that better German knowledge and lower perceived difficulty of the task are positively associated with speed of typing.

Table 7

Correlations between output parameters and task proficiency variables in the control time

	Task easiness	German knowledge
Preciseness_10	,015	**,208**[*]
Mistakes_10	**-,298**[**]	**-,375**[**]
Speed_Minutes_10	**,229**[*]	,122
Speed_Words_10	**,328**[**]	**,398**[**]

*Correlation is significant at the 0.05 level (2-tailed).

**Correlation is significant at the 0.01 level (2-tailed).

Another variable that significantly correlates with the "task proficiency" variables is the number of mistakes in the transcript meaning that higher task proficiency can be related to the lower number of mistakes. Following our suggestion that preciseness and speed of transcribing are variables based on their situational importance for individual, we see that each of them has a weak correlation with only one of the two proficiency variables – preciseness with language knowledge variable (r=,208, p<0,05) and speed of transcribing with perceived easiness of the task (r=,229, p<0,05).[46]

Table 8

Correlations among output parameters in the main time

	Speed_Minutes	Preciseness	Mistakes
Speed_Words	**,688**[**]	**,374**[**]	**-,324**[**]
Speed_Minutes		**-,216**[*]	**-,213**[*]
Preciseness			**-,274**[**]

*Correlation is significant at the 0.05 level (2-tailed).

**Correlation is significant at the 0.01 level (2-tailed).

As we saw in the Table 6 in the control time parameters of the output quantity were strongly interrelated, which was also the case in the main time (Table 8). The correlation had the same magnitude and we can say that the common variance composed about 50 percent. We can also observe that there is still a significant negative correlation between speed of transcribing and preciseness, however the strength of correlation has decreased and instead of more than 35 percent of common variance in the control time we can speak of only less than 5 percent in the main time.

At the same time speed of typing correlates positively in the main time with preciseness and negatively with amount of mistakes (both correlations are significant at p<0,01). This contradicts the fact that higher accuracy of typing can be associated with lower speed of typing. But if we also take into account that two quality parameters also correlate negatively (higher preciseness is associated with higher accuracy of typing), we can suggest that those participants who were better at quality were also better in quantity in terms of typing speed. In general this can mean that those participants who had higher typing skills could also provide better quality of output. At the same time such line of argument can also mean that those participants with lower typing skills were providing lower quality. This is partly true, but the correlation table shows that lower speed of transcribing is associated with higher number of mistakes, but also with higher preciseness. So, there is a trade-off between speed of transcribing and preciseness, but not between speed of transcribing and accuracy of typing.

---

[46]Correlation between the variables Task easiness and German Knowledge is 0,293 (p=0,04)

Thus, summing up the results of the above analysis we can conclude that:

- In the control time speed of typing strongly correlates with "task proficiency" variables and it is positively correlated with speed of transcribing.

- Preciseness of transcribing within control time reflects the perceived importance of preciseness and it is negatively correlated with speed of transcribing, meaning that there is a trade-off between preciseness of the transcript and speed of transcribing (or maximizing the amount of the recording transcribed).

- Number of mistakes in the control time positively correlates with task proficiency variables and with preciseness, which can mean that it is an issue of choice restricted by task proficiency.

- Correlations between output variables in the main time reflect the trade-off between preciseness and speed of transcribing. At the same time the pattern of correlations implies that those employees who provided higher quantity of output in terms of amount typed were also often the ones with higher preciseness and lower amount of mistakes.

5.4.2. Analysis of the parameters within groups and comparison between groups

In order to understand the pattern and essence of changes in output variables it is useful first to throw another glance at the relationships between them. As we have already discussed the relationship between the output parameters in the main time for the whole sample, now we proceed with the analysis of these relationships within each treatment group.

We start by exploring the relationship between the output variables in the control and the main time. It could seem natural to believe that higher level of output variable in the control time will be associated with the comparatively higher level of the same variable in the main time. However, as we see from the Table 9 it is not always the case. In particular, analysing the correlations between, for example, preciseness in the control time and in the main time for each treatment group we can see that in the groups with the requirement of non-anonymity of the employee there is indeed a significant relationship between initial preciseness and preciseness in the main time (for Group 3 $r=0,5$, $p<0,05$ and for Group 4 $r=0,6$, $p<0,01$). At the same time for two groups with anonymous employees the relationship, though also positive, is non-significant (for both groups $p>0,05$), which means that lower initial preciseness can be associated with higher preciseness in the main time and the other way round. At the maximum initial preciseness explains about 30 percent of variation in the preciseness in the main time. This leaves a lot of room for other factors to influence this variable and supports our initial proposition, stating that preciseness is a variable determined by the intentional choice based on the beliefs about its importance.

Amount of mistakes, on the other hand, according to the Table 9 is the output variable with the strongest correlations between its levels in the control and the main time. For three out of four groups

amount of mistakes in the control time shared about 65 percent of variation with the amount of mistakes in the main time.[47]

Speed of transcribing in the main time is also strongly related to the initial speed of transcribing, but the correlations are much weaker in the non-anonymous than in anonymous groups (in Group 4 it is even not significant at p<0,05).

Finally for speed of typing we observe the situation where the variation in the initial parameter explains from about 25 percent of variation in the main time for Group 3 and up to 65 percent of variation in Group 1. At the aggregated level we can suggest that initial speed of typing shares more variation with the parameter of typing speed in the main time in the anonymous groups and leaves more room for other influences in the non-anonymous ones.

Table 9

Correlations between output parameters in the control and the main time

Group	Preciseness	Mistakes	Speed_Minutes	Speed_Words
1	,283	**,822**[**]	**,670**[**]	**,823**[**]
2	,195	**,570**[**]	**,625**[**]	**,722**[**]
3	**,505**[*]	**,801**[**]	**,560**[**]	**,543**[**]
4	**,574**[**]	**,827**[**]	,395	**,670**[**]

*Correlation is significant at the 0.05 level (2-tailed).

**Correlation is significant at the 0.01 level (2-tailed).

Further we discuss the correlations among the output variables in the main time (Table 10). As with the aggregated level data, also in the individual groups there is a pattern of significant correlation between parameters of the output quantity (speed of transcribing and speed of typing) for all groups, but the strongest this relationship is in Group 1. In fact more than 80 percent of the variation in one parameter of the output quantity is shared with the other, while none of the other parameters are correlated in this group.

Those participants in Group 1 who possessed better typing skills typed faster and achieved higher speed of transcribing. However, this was not achieved by reducing preciseness, as follows from a negligible size of correlation coefficient for preciseness and speed of transcribing (r=-0,049, p>0,1).

---

[47] Due to the design of the experiment we can argue that here we have a unidirected influence, as only behavior in the control time could be a predictor of the behavior in the main time and not the other way round.

Table 10

Correlations among output parameters in the main time (group split)

Group		Mistakes	Speed_Minutes	Speed_Words
1	Preciseness	-,240	-,049	,164
	Mistakes		-,220	-,247
	Speed_Minutes			**,951**[**]
2	Preciseness	-,197	**-,428**[*]	,138
	Mistakes		-,088	-,363
	Speed_Minutes			**,731**[**]
3	Preciseness	-,244	-,316	**,595**[**]
	Mistakes		-,113	-,090
	Speed_Minutes			**,416**[*]
4	Preciseness	-,366	-,071	**,600**[**]
	Mistakes		**-,439**[*]	**-,703**[**]
	Speed_Minutes			**,669**[**]

*Correlation is significant at the 0.05 level (2-tailed).

**Correlation is significant at the 0.01 level (2-tailed).

Thus, in Group 1 participants' effort to maximise the amount of recording transcribed was not influenced by the quantity-quality trade-off, but depended for example on the task proficiency of the participants.

The trade-off between speed of transcribing and preciseness is clearly observed in Group 2, where participants with high speed of transcribing were the ones with low preciseness and vice versa (r=0,43, p<0,05). Group 2 is the only group with significant negative correlation between these variables.

As follows from the previous analysis getting the additional information about particular customer in Group 2 led to the much less unified focus on prescribed standard of preciseness. Instead participants focused on maximisation of amount transcribed.

In both Groups 1 and 2 we do not find the significance for the pattern observed in the aggregated sample related to division between participants according to their task proficiency. However, this pattern becomes significant in Group 3 and especially in Group 4. In Group 3 there is a clear link between preciseness and speed of typing, while in Group 4 higher speed of typing means also lower amount of mistakes.

Thus, in both groups where the name of employee is opened to the customer, higher task proficiency is associated with higher quality. However, in Group 3 better accuracy is not necessarily the prerogative of highly skilled.

In Group 4 where two treatment effects were combined participants had to think simultaneously about the customer and their own image. If we combine the conclusion regarding the motivation in previous groups we can suggest that participants in Group 4 wanted to provide high level of both quality

parameters (preciseness as Group 1 and accuracy as Group 3), while achieving the highest speed of transcribing possible (as Group 2).

In this case, if they were united in their actions, the only difference should be in task proficiency. And this is what we see in the data: those with better quantity were also the ones with better quality, while participants with lower quality of transcripts also provided not more but less of a transcript.

5.5. Change in focus from the control time to the main time

5.5.1. Analysis of the parameters for the whole sample

Already in the descriptive analysis of the main output variables in the main time through their comparison with the respective values in the control time we could observe that the mean (and the median) level of all variables had changed. One of the most important questions in our discussion is whether the magnitude of these changes is large enough to talk about statistical significance. To answer this question we conduct both non-parametric and parametric related samples analysis, allowing us to compare the level of output variables reached in the control time with the respective parameters in the main time. First to assess the general trends we analyse such changes using the data for the whole sample. Afterwards we split the sample according to the treatment information provided to the participants and conduct separate related samples analysis for each group. This allows us to see whether participants in the groups focused on the same or different parameters of output.

If we take the whole sample for analysis without division for treatment groups we discover that there are two output variables that show clear change pattern: preciseness and accuracy (Figure 7).

In particular, the results of the related-samples Wilcoxon Signed Rank Test show that the preciseness grew significantly ($z=2,054$, $p<0,05$) and the number of mistakes decreased even more significantly ($z=-3,232$, $p<0,01$). While 60 out of 100 participants decreased the amount of mistakes, only 40 increased this amount. Also as one can see from Figure 7 among those who increased the number of mistakes this increase was very small (more than 30 of 40 participants increased mistakes only by not more than 2,5 mistakes per 100 words). At the same time among those who decreased mistakes more than a half decreased them by a larger amount - from 2,5 to 20 mistakes per 100 words.

Figure 7

Change in Preciseness and Number of Mistakes (main time - control time)
(Related samples Wilcoxon Signed Rank Test)

Preciseness	**Number of mistakes**

Total N	100
Test Statistic	3.122,500
Standard Error	290,839
Standardized Test Statistic	2,054
Asymptotic Sig. (2-sided test)	,040

Total N	100
Test Statistic	1.585,000
Standard Error	290,839
Standardized Test Statistic	-3,232
Asymptotic Sig. (2-sided test)	,001
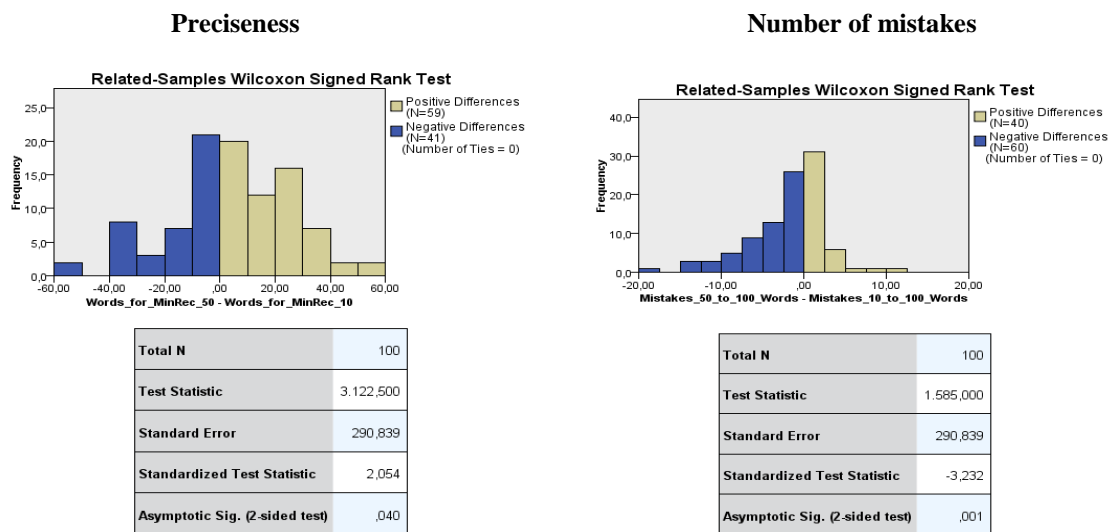
These observations are also supported by the results of parametric analysis, which shows significant decrease in mistakes at the aggregated level and (though less significant) increase in preciseness. Interestingly the parametric related-samples comparison also shows significant negative difference (and confidence interval not including zero) of speed of transcribing in the control and the main time. This means that general improvement of quality of output came together with slight decrease in speed of transcribing, but did not significantly influence the speed of typing.

Table 11

Change in output parameters (control time – main time)
(Related samples ANOVA, bootstrapped)[48]

Parameter	Mean	Bootstrap				
		Bias	Std. Error	Sig. (2-tailed)	BCa 95% CI	
					Lower	Upper
Preciseness	-4,245	,128	2,212	,056	-8,696	,172
Mistakes	1,844	-,026	,462	,001	,923	2,764
Speed_Minutes	,034	-,001	,0158	,055	,008	,063
Speed_Words	,340	,001	,634	,598	-,789	1,631

One of the reasons for the lack of significant change in the parameters of the output quantity across the sample can be related to the counter-directional differences among the groups. If participants, as we

---

[48] Here and further unless otherwise noted, bootstrap results are based on 1000 bootstrap samples.

argue, assigned different situational importance to different output parameters due to the treatment information, then, if statistically they were moving in significantly different directions, by aggregation these differences can cancel out. Indeed, significant improvement of the preciseness could result in the decrease of speed of transcribing by those participants who attempt significant increase in preciseness, while others keeping or decreasing preciseness could win the speed. Dramatic decrease in mistakes in typing can lead to slowing down the speed of typing, but less significant changes can allow keeping the speed or even increasing it. That is why by seeing the significant increase of quality we could not necessarily expect decrease in quantity.

Thus, the next step of the analysis includes the comparisons between output variables in the control and main time for each treatment group separately.

5.5.2. Analysis of the parameters within groups and comparisons between groups

Preciseness

We start with the parameter of preciseness (Table 12) and see that all groups increased preciseness in the main time (test statistic for all of them is based on negative ranks), but the only group where the increase in preciseness was significant was Group 1 ($z=-2,16$, $p=0,03$).

Table 12

Groups split for change in Preciseness (main time – control time)
(Wilcoxon Signed Ranks Test Statistics)

Group		Statistic
1	Z	**-2,162[a]**
	Exact Sig. (2-tailed)	**,030**
2	Z	-,955[a]
	Exact Sig. (2-tailed)	,353
3	Z	-,600[a]
	Exact Sig. (2-tailed)	,565
4	Z	-,429[a]
	Exact Sig. (2-tailed)	,684

a. Based on negative ranks.

This is also supported by the data in the Table 13, where besides observing the mean ranks we can also see how many participants changed preciseness in positive versus negative direction. Indeed, in Group 1 only 8 people decreased preciseness while 19 participants increased it.

In Group 2 and Group 4 (groups with the information about a selected customer) the results are balanced in numbers: 14 participants increased preciseness and 11 and 10 participants respectively decreased it.

Table 13

**Groups split of ranked scores for change in Preciseness
(main time – control time)**

Group	Parameter	N	Mean Rank
1	Negative differences	8[a]	12,38
	Positive differences	19[b]	14,68
	Total	27	
2	Negative differences	11[a]	11,55
	Positive differences s	14[b]	14,14
	Total	25	
3	Negative differences	12[a]	10,75
	Positive differences	12[b]	14,25
	Total	24	
4	Negative differences	10[a]	13,50
	Positive differences	14[b]	11,79
	Total	24	

a. Preciseness_50 < Preciseness _10

b. Preciseness_50 > Preciseness _10

In Group 3 the number of those who increased and decreased preciseness was completely the same (12 participants). While the absolute difference between mean ranks is the highest in this group, the fact that the participant there were equally willing to increase and decrease preciseness makes the overall result insignificant.

The significant improvement of preciseness by Group 1 is partly supported by the results of parametric test (Table 14).

Table 14

Groups split for change in Preciseness (control time - main time)
(Related samples ANOVA, bootstrapped)

Group	Mean	Bootstrap			BCa 95% CI	
		Bias	Std. Error	Sig. (2-tailed)	Lower	Upper
1	-7,201	,064	3,545	,055	-13,883	,056
2	-3,652	-,132	3,495	,338	-9,894	2,372
3	-3,947	-,167	5,705	,486	-14,725	6,529
4	-1,833	-,237	4,510	,713	-10,846	6,325

Remembering that Group 1 was the only group with significant increase in the preciseness of transcript, we can make a conclusion that general information about the purpose of the task combined with explicit requirement for the preciseness of the transcript moved the focus (and with it the relative

situational importance) in the direction of preciseness. Most of the participants regardless of their initial typing abilities concentrated on providing more precise transcript.

While other groups do not show significant increase in preciseness analysed on their own, we check whether information about the employee or the customer was important on a more aggregated level.

Here and also for the other main output variables we combine Groups 3 and 4 (groups with the information about the employee required) and compare the results with the remaining Groups 1 and 2 (without such information required) to see, whether there is a difference in results triggered by requiring the information about the employee. For statistical analysis we create the dummy variable for employee's non-anonymity and code the participants from Groups 3 and 4 as "1", and from Groups 1 and 2 as "0".

Similarly we combine Groups 2 and 4 (groups with the information about the selected customer) and compare the results with the remaining Groups 1 and 3 (without such information) to see, whether there is a difference in results triggered by providing the information about the selected customer. For statistical analysis we create the dummy variable for customer's non-anonymity and code the participants from Groups 2 and 4 as "1", and from Groups 1 and 3 as "0".[49]

Figure 8

Change in Preciseness (main time - control time)
(Related samples Wilcoxon Signed Rank Test)

Employee_NonAnonymous   **Employee_Anonymous**
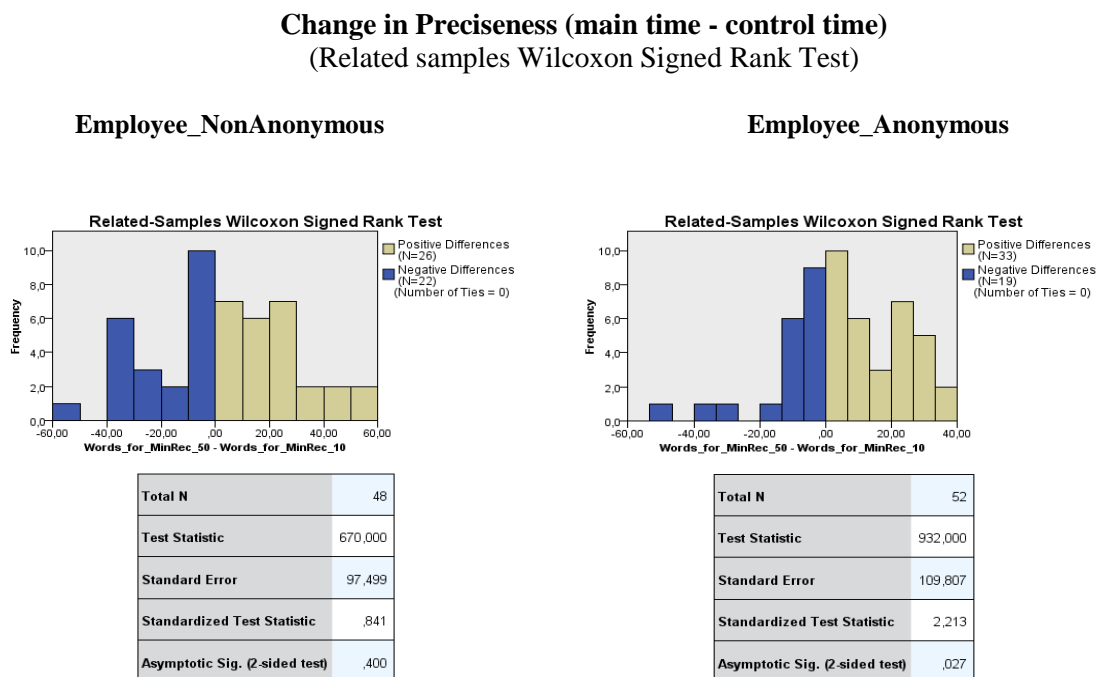
Figure 8 shows that Groups 1 and 2 - having no requirement to open employee's identity - significantly increased preciseness ($p<0.05$), while Groups 3 and 4 (without such requirement) did not do it ($p>0.1$).

---

[49] For the parameter of preciseness the comparisons made using the dummies for customer's non-anonymity provided less clear-cut results compared to the employee's non-anonymity.

This analysis brings us to the conclusion that upon receiving general information about the purpose of the task participants significantly increased the importance of preciseness, but additional information - either opening the customer's identity or requiring to provide employee's own name - made the increase in preciseness insignificant. This means that additional information provided within treatment groups moved the focus to other parameters of output.

At the aggregated level comparing groups with and without employee's non-anonymity we see that only the employees who were not required to open their identity significantly increased preciseness, while non-anonymous employees were less united about such an increase. We believe that the necessity to open identity moved focus to other attributes of the task and decreased focus on preciseness.

Number of Mistakes

Next parameter we focus on in our analysis is the change in the number of mistakes participants made in the main time compared to the control time. In the Table 15 we see that all groups reduced the amount of mistakes in the text (all statistics is based on positive ranks), but the only group which reached significance in this change was Group 3 with $z=2,857$, $p<0.01$. Group 2 on the contrary had the smallest improvement in the number of mistakes among all groups.

Table 15

Groups split for change in Number of Mistakes (main time – control time)
(Wilcoxon Signed Ranks Test Statistics)

Group	Statistic	Difference
1	Z	-1,514[a]
	Exact Sig. (2-tailed)	,135
2	Z	-,659[a]
	Exact Sig. (2-tailed)	,525
3	Z	-2,857[a]
	Exact Sig. (2-tailed)	**,003**
4	Z	-1,543[a]
	Exact Sig. (2-tailed)	,128

a. Based on positive ranks.

Table 16 shows that in Group 3 only 7 people increased the number of mistakes, while 17 people decreased this number. Similar situation can be observed in Group 4 where 15 out of 24 participants decreased the number of mistakes.

Table 16

Groups split of ranked scores for change in Number of Mistakes
(main time - control time)

Group	Parameter	N	Mean
1	Negative differences	16[a]	15,75
	Positive differences	11[b]	11,45
	Total	27	
2	Negative differences	12[a]	15,58
	Positive differences	13[b]	10,62
	Total	25	
3	Negative differences	17[a]	14,71
	Positive differences	7[b]	7,14
	Total	24	
4	Negative differences	15[a]	13,60
	Positive differences	9[b]	10,67
	Total	24	

a. Mistakes_50_to_100_Words < Mistakes_10_to_100_Words

b. Mistakes_50_to_100_Words > Mistakes_10_to_100_Words

In the groups without the requirement of opening the employee's identity the situation is much less clear: while in Group 1 there are more people decreasing than increasing the amount of mistakes, in Group 2 the numbers are almost equal and it is the only group where technically number of people improving accuracy is lower than the number of those increasing the number of mistakes.

The general difference in the mean ranks in Group 3 is obviously the biggest among groups. For this group we can observe the situation where even those who increased the number of mistakes increased it for not more than 2,5 words to 100 words of typed text, while those who reduced the number of mistakes operated within the range of up to 15 mistakes per 100 words. However, Groups 1 and 4 also show quite large difference between both numbers of people changing the number of mistakes in opposite directions and also between mean ranks (though this difference is not enough for significant result)..

Parametric analysis also shows significant improvement in the amount of mistakes for all groups (except for Group 2), but the highest change was introduced by the participants in Groups 3 and 4 (Table 17).

Table 17

Groups split for change in Number of Mistakes (control time - main time)
(Related samples ANOVA, bootstrapped)

| Group | Mean | Bootstrap | | | BCa 95% CI | |
		Bias	Std. Error	Sig. (2-tailed)	Lower	Upper
1	1,395	,025	,794	,083	-,162	3,122
2	,719	-,048	,928	,456	-1,032	2,392
3	3,689	-,062	1,106	,018	1,759	5,862
4	1,675	-,023	,759	,047	,318	3,058

It is also important that the means for Groups 3 and 4 have strictly positive confidence intervals, while the confidence intervals for Groups 1 and 2 include zero.

Aggregating the results for Groups 3 and 4 under the dummy for employee's non-anonymity and conducting non-parametric related samples Wilcoxon Signed Rank Test (Figure 9) we get the strong support for the statistically significant decrease in the number of mistakes for groups where employees were non-anonymous (while the results are insignificant for groups with anonymous employees).

Figure 9

Change in Number of Mistakes (main time - control time)
(Related samples Wilcoxon Signed Rank Test)
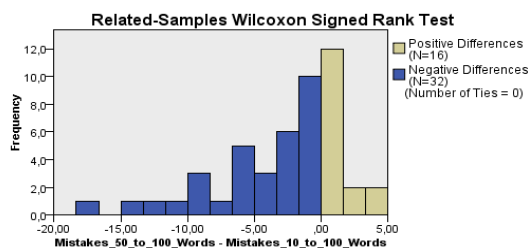
Employee_NonAnonymous   **Employee_Anonymous**
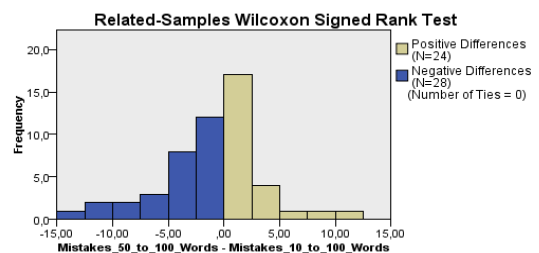
From this analysis we can make the following conclusions: while all groups reduced the number of mistakes after getting the treatment information, participants of the groups required to open their identity (especially of Group 3) had more agreement on this issue. Introduction of the requirement of employee's non-anonymity led to the higher individual visibility of the participants' output in Groups 3 and 4, which significantly changed the relative situational importance of accuracy (which can be seen as a highly visible non-prescribed parameter of quality). Other groups were less united in this attempt.

Speed of Transcribing

As was shown in the previous analysis both parameters of quantity did not significantly change on aggregated level, but, as we argued, this could be a result of counterdirectional changes in the groups; for example, if one group significantly increased the parameter and another decreased it, the aggregated result could lack any significant support for change in one direction. Indeed, our further analysis shows that at the level of individual groups the patterns of behavior are far from being unified.

On the one hand the parameter of speed of transcribing did not significantly change in the main time compared to the control time within groups (Table 18).

Table 18

Groups split for change in Speed of Transcribing (main time – control time)
(Wilcoxon Signed Ranks Test Statistics)

Group		Statistics
1	Z	-,601[a]
	Exact Sig. (2-tailed)	,562
2	Z	-1,466[b]
	Exact Sig. (2-tailed)	,148
3	Z	-1,629[a]
	Exact Sig. (2-tailed)	,107
4	Z	-1,029[a]
	Exact Sig. (2-tailed)	,317

a. Based on negative ranks
b. Based on positive ranks

However, we still can see that all groups except Group 2 decreased the speed of transcribing. Group 2 is the only one which increased this parameter, though insignificantly (z=-1,47, p=0,148).

At the same time there is a clear difference between those who increase and decrease the speed of transcribing in Group 2 and Group 3 (Table 19). Interestingly in Group 2 we observe that the majority of participants (16 out of 25) increased the speed of transcribing, while in Group 3 almost the same

proportion of participants (15 out of 24) decreased this parameter. In Groups 1 and 4 there was almost equal amount of participants, who changed the speed of transcribing in one or another direction.

**Groups split of ranked scores for change
in Speed of Transcribing (main time - control time)**

Group	Parameters	N	Mean Rank
1	Negative differences	13[a]	16,46
	Positive differences	14[b]	11,71
	Total	27	
2	Negative differences	9a	12,00
	Positive differences	16b	13,56
	Total	25	
3	Negative differences	15a	13,80
	Positive differences	9b	10,33
	Total	24	
4	Negative differences	12a	15,50
	Positive differences	12b	9,50
	Total	24	

a. Speed_Minutes_50 < Speed_Minutes_10

b. Speed_Minutes_50 > Speed_Minutes_10

Results of parametric related samples comparisons also do not show any clearly significant changes in the speed of transcribing for separate groups.

Groups split for the change in Speed of Transcribing (control time - main time)
(Related samples ANOVA, bootstrapped)

Group	Mean	Bootstrap				
		Bias	Std. Error	Sig. (2-tailed)	BCa 95% CI	
					Lower	Upper
1	,020	-,0010	,0157	,234	-,0066	,0446
2	-,024	,0004	,0150	,119	-,0541	,0051
3	,092	-,0000	,0537	,183	-,0076	,2049
4	,054	-,0014	,0336	,147	-,0045	,1175

The combined analysis for groups with and without the requirement for non-anonymity of the employee (Figure 10) supported the expectation of the decreasing trend of speed of transcribing for the groups with the requirement of provision of employee's name. Here we see that groups required to open identity were more often decreasing the speed and less often increasing it (z=-1,969, p<.05). For

participants from the groups without requirement to open identity the change in the speed of transcribing was not significant.

Change in Speed of Transcribing (main time - control time)
(Related samples Wilcoxon Signed Rank Test)

Employee_NonAnonymous   **Employee_Anonymous**
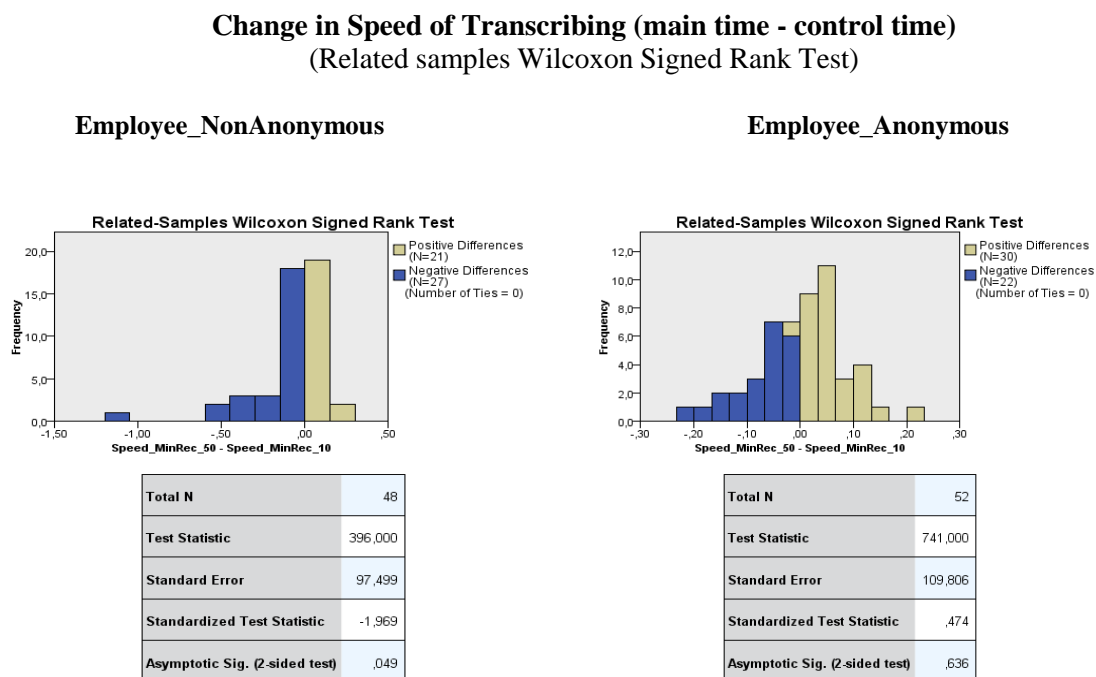
The conclusion for the analysis of the change between speed of transcribing in the main time compared to the control time is the following: while the results for the individual groups do not show a clear change in the speed of transcribing, at the more aggregated level groups with the non-anonymity requirement significantly decreased this parameter. In particular, Group 2 showed a pattern of behavior different from other groups, namely its participants were more likely to increase speed of transcribing than to decrease it.

Speed of Typing

Last parameter of output which we are going to analyse is the speed of typing. Comparing the speed of typing in the main time to the control time (Table 21) we see that Groups 1 and 2 increased the speed of typing (statistic is based on negative ranks) and Groups 3 and 4 decreased this parameter (statistic is based on positive ranks).

Moreover for two groups – Group 2 and Group 3 – the change in speed of typing was significant at $p<0,1$ level. While participants in Group 2 significantly increased the speed, the participants from Group 3 decreased it.

Table 21

Groups split for change in Speed of typing (main time – control time)
(Wilcoxon Signed Ranks Test Statistics)

Group	Parameter	Statistic
1	Z	-1,394[a]
	Exact Sig. (2-tailed)	,168
2	Z	-1,749[a]
	Exact Sig. (2-tailed)	,081
3	Z	-1,900[b]
	Exact Sig. (2-tailed)	,057
4	Z	-,557[b]
	Exact Sig. (2-tailed)	,589

a. Based on negative ranks.

b. Based on positive ranks.

If we look at the number of participants in each group who increased or decreased the speed of typing (Table 22) we observe that in Group 2 (with information about the selected customer) 16 out of 25 participants increased the speed of typing and in Group 3 (with non-anonymity of the employee) 14 out of 24 participants decreased the speed of typing. In Group 1 there was clearly higher proportion of employees increasing the speed of typing than those decreasing it, while in Group 4 the difference was almost negligible.

Table 22

Groups split for the ranked scores for change in Speed of Typing
(main time – control time)

Group	Parameter	N	Mean Rank
1	Negative Ranks	11[a]	11,91
	Positive Ranks	16[b]	15,44
	Total	27	
2	Negative Ranks	9[a]	10,83
	Positive Ranks	16[b]	14,22
	Total	25	
3	Negative Ranks	14[a]	15,46
	Positive Ranks	10[b]	8,35
	Total	24	
4	Negative Ranks	13[a]	13,04
	Positive Ranks	11[b]	11,86
	Total	24	

a. Speed_typing_50 < Speed_ typing _10

b. Speed_ typing _50 > Speed_ typing _10

The parametric test (related samples ANOVA) also renders support for the results shown above (Table 23). In particular, Groups 1 and 2 indeed have a positive mean difference in the speed of typing, while Groups 3 and 4 have a negative difference. Though the significance for means difference in the Groups 2 and 3 lies at the level of $p<0,1$, confidence intervals (not including zero values) reflect the general pattern of increase in the typing speed for Group 2 and decrease in in this parameter for Group 3.

<div align="right">Table 23</div>

Groups split for change in Speed of Typing (control time - main time)
(Related samples ANOVA, bootstrapped)

Treatment	Mean	Bootstrap				
		Bias	Std. Error	Sig. (2-tailed)	BCa 95% CI	
					Lower	Upper
1	-1,219	-,009	,807	,158	-2,882	,402
2	-1,886	,015	,897	,062	-4,010	-,207
3	4,122	-,011	1,914	,081	,793	8,084
4	,631	,006	1,037	,552	-1,386	2,507

These results gain more significance if we combine groups with and without requirement to open employee's identity (Figure 11). In the groups without such requirement (Groups 1 and 2) we can observe an increase in the speed of typing with higher significance as in Group 2 alone ($p<0.05$). there is also a slight increase in the significance for the groups with the requirement of employee's non-anonymity (Groups 3 and 4) who decreased the typing speed ($p<0,1$).

This means first of all that in our task employees were generally able to increase the speed of typing. Although, as we argued in the beginning, this parameter was restricted by the proficiency of the participants, it still had a potential to grow. By requiring the employees to provide their names we ended up with the decrease in speed of typing. We know that groups in this category were not significantly increasing preciseness, but instead focused on accuracy. Thus, the significant increase in accuracy was "too costly" for Groups 3 and 4, as it led to decrease in typing speed. Participants from groups without the requirement to open their names were less united in their focus: Group 1 significantly increased preciseness and also slightly increased accuracy, but Group 2 did not significantly increase any of the parameters related to the quality of output. However, Group 2 managed to become the only group which achieved a slight increase in the speed of transcribing and a more significant increase in the speed of typing.

Figure 11

Change in Speed of Typing (main time - control time)
(Related samples Wilcoxon Signed Rank Test)
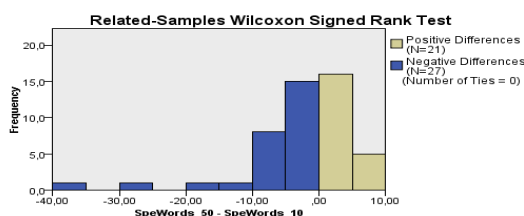
Employee_NonAnonymous	Employee_Anonymous
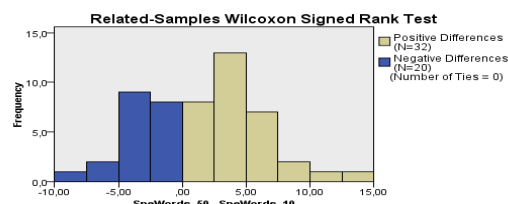
Concluding this part we can sum up the results discussed above for each of the treatment groups:

- Group 1. Taking into account the fact that the participants of Group 1 significantly increased preciseness, we consider the observations that they managed not to decrease the parameters of quantity significantly (at the same time not deteriorating in the accuracy) as a sign of general increase in effort which apparently led to the increase in efficiency[50].

- Group 2. This group did not improve significantly any of the parameters of output quality, but at the same time it did not deteriorate the quantity of output. In fact this group was the only one where participants significantly increased output quantity without decreasing its quality, which leads us to interpreting these results in terms of significant change in focus (compared to the control time) in the direction of quantity maximization for this group as well as in terms of the general increase in effort.

- Group 3. While having significantly increased the accuracy of transcript, this group was less unified regarding the importance of the preciseness. Earlier analysis showed that the accuracy was not significantly correlated with the speed of typing, which means that the number of mistakes was reduced by both more and less skilled participants. .At the same time a significant increase in accuracy (decrease in the number of mistakes) led to the significant decrease in the speed of transcribing and in speed of typing.

- Group 4. This group improved both qualitative parameters but did not reach the threshold of significance at any of them. It was either not united regarding the relative importance of the

---

[50] We admit that such increase could have been reached due to the learning effect. However we assume that learning itself requires effort and therefore should be motivated. That is the reason why we assume the increase in general motivation die to the treatment effect, which in turn could lead to higher persistency and/or higher efficiency.

selected parameters or its participants had a more balanced approach towards such focus and improved all parameters simultaneously. Generally it is not counterintuitive to suggest that this group can be characterized by mixed influence of treatment effects.

5.6. Questionnaire-based variables and their relationship with the main output parameters

5.6.1. Descriptive statistic for the questionnaire-based variables

As was explained in the chapter "Method of research" the design of our study involved the behavioral experiment which was complimented by a number of questionnaires. As a result we were able to obtain the individual-level data for several additional variables which we used for the analysis. We split these variables into two major parts

- variables related to the demographics of the sample and individual task proficiency;
- variables related to the personal predispositions for empathic trait, agreement with descriptive and prescriptive norms, social value orientation as well as variables reflecting situation-related emotions like empathic and distress states.

We have already descriptively discussed the first type of variables in the beginning of this chapter. Here we will focus on the variables related to hypotheses testing and further discuss the relationship between them and variables related to the task proficiency.

The descriptive statistic for variables reflecting personal predispositions and states is shown in the Table 24.

Table 24

Descriptive statistic for the variables reflecting personal predispositions and states

		Empathic trait	Prescriptive norm	Descriptive norm	Empathic state	Distress state
N	Valid	100	98	98	98	98
	Missing	0	2	2	2	2
Mean		12,25	5,90	4,38	3,85	2,89
Median		12,50	6,00	5,00	3,93	2,79
Std. Deviation		2,52	1,21	1,33	1,15	1,15
Minimum		5,0	1	1	1	1
Maximum		16,0	7	7	6,57	5,86

Table 24 shows that the average value for the empathy measured as trait lies at the level of 12,25 points (with minimum at 5 points and maximum at 16 points).

Both questions regarding the agreement with the descriptive and prescriptive norms were answered using the 7-points Likert scale. We can see that participants ranged from 1 ("not at all") to 7 ("very much") in their level of agreements with the statements regarding normative and observed behavior of others towards people with limited abilities. On average they rather strongly agreed with the prescriptive norm of helping (with average score of around 6 points), but were more sceptical about real behavior of people (average score of about 4,4 points).

Interestingly (and importantly for our future analysis) almost 70 percent of participants strongly believed that it is necessary to help people with disabilities (scores of 6 and 7) and only 5 percent did not share this opinion (scores from 1 to 3). As for the descriptive norm, only 22 percent of participants believed that other people really help people with disabilities (scores of 6 and 7) and almost 30 percent thought that this is not the case.

Thus, we observe a clear difference in what people consider as "should be done" and what "is done in reality". Many participants believe that while it is necessary to help, in reality people do not do it.

Upon completion of the experiment we also measured the situational empathy and distress and found out that on average participants did not feel particularly empathic (mean value around 4 points), but importantly even less they felt distressed (mean value around 3 points).

5.6.2. Relationship between questionnaire-based variables and main output parameters

Further we provide the analysis of relationships between each of the output parameters and a number of descriptive variables measured on the basis of questionnaires provided before and after the experimental session.

It is important to mention that in the tables below for the sake of clarity and simplicity we show not all of the parameters but only those which showed significant correlation (Spearmen's Rho) with the output variable in at least one of the treatments.[51] Thus, the absence of a descriptive variable in a particular correlation table means that this variable did not correlate with the output variable of interest in any of the treatments and was omitted from the table.

Preciseness
First we look at the parameter of preciseness and its relationship with questionnaire-based variables. Preciseness is positively correlated with language proficiency for all groups, however only in Groups 1 and 4 this correlation is strongly significant (r=0,53, p<0,01 for Group 1 and r=0,67, p<0,01 for

---

[51] The figures for the correlations which are significant at p<0,1 are shown in the tables in **bold**.

Group 4). This renders additional support to the fact that in both of these groups quality in terms of preciseness is not an issue of trade-off, but the parameter, which importance is accepted by all group members and ability to maximise it depends on the general abilities in task performance of the participant. In the Groups 2 and 3 language proficiency ceases to be an important issue for the discussion of preciseness.

As for the relationship between preciseness and variables measuring the construct of empathy and norm adherence, the only group where higher preciseness was associated with higher empathy as trait and lower agreement with descriptive norm statement was Group 4. In other words in this group those participants who were more empathic and believed that others do not help people with limited abilities made more precise transcripts.

Table 25

Groups split for correlations of
Preciseness with selected questionnaire-based variables

Group	Empathic trait	Prescriptive norm	Descriptive norm	German knowledge	Task interest
1	-,105	**-,385**[*]	,348	**,527**[**]	-,201
2	,155	-,126	**,508**[*]	,069	**-,430**
3	,340	-,249	,187	,253	-,060
4	**,451**[*]	-,072	**-,411**[*]	**,669**[**]	-,049

*Correlation is significant at the 0.05 level (2-tailed).

**Correlation is significant at the 0.01 level (2-tailed).

In order to look deeper into the relationships between preciseness and empathic trait we checked whether non-linearity could be an issue. A comparative table below gives two models for preciseness as the function of empathic trait, with two types of functions – linear and quadratic (Table 26). In Group 2 the relationship is indeed better described not by linear relationship, but by quadratic function, which explains about 30 percent of the variation in preciseness. In Group 4 linear function explains about 15 percent of variance.

Table 26

Preciseness as a function of Empathic trait
Groups split for comparison of the linear and quadratic regression models

Group	Equation	Model Summary					Parameter Estimates		
		R Square	F	df1	df2	Sig.	Constant	b1	b2
2	Linear	,078	1,934	1	23	,178	39,623	3,545	
	Quadratic	**,291**	4,513	2	22	**,023**	-495,315	89,144	-3,358
4	Linear	**,151**	3,921	1	22	**,060**	26,434	4,544	
	Quadratic	,155	1,931	2	21	,170	69,075	-2,432	,279

Relationship between the preciseness of the transcript and the level of agreement with the statement, that usually people help other people with disabilities (descriptive norm) is significant in both Group 2 and 4. However, in Group 2 correlation is positive and in Group 4 it is negative.

If we remember that the participants provided with the information about individual customer (Group 2) were focused on the quantity of output, we can suggest that they believed that they were solely responsible for the amount of output provided to the customer. In other words the information provided to the participants in Group 2 (information about customer) led them to believe that none of the other participants transcribed or will further transcribe the recording that they were transcribing. That is why they focused on the increase of the amount transcribed/typed.

The positive correlation between agreement with the statement "other people help disabled people" and preciseness can be interpreted as a belief that others will also help in this task and contribute to the output. Thus, participants with high score on this variable were more focused on preciseness and less on the amount transcribed, compared to the participants having doubts about others' behavior.[52].

We checked the significance of the relationship between preciseness and agreement with the descriptive norm by calculating the function with preciseness as dependent variable and descriptive norm as independent variable. As a result in the absence of the initial preciseness (in the control time) as covariate, agreement with the descriptive norm of helping explains more than 40 percent of variation in preciseness of the transcript in the main time using the linear model (and up to 66 percent in the quadratic model). However, this connection exists only in Group 2 (where the employee is provided with the information about the customer). As soon as the employee becomes non-anonymous (Groups 3 and 4) focus moves to other attributes of output and significance of the relationship disappears.

Number of Mistakes

When we test the relationships between the number of mistakes that participants made in the main time and the questionnaire-based variables (Table 27) we observe in all groups, but especially in Groups 2 and 4, the significant negative correlations between the number of mistakes and variables related to abilities – easiness of the task and language knowledge.

---

[52] As we will see further this is supported by the significant negative correlation between speed of transcribing and agreement with descriptive norm (statement about other's helping behavior).

Table 27

**Groups split for correlations
of Number of Mistakes with selected questionnaire-based variables**

Group	Empathic trait	Empathic state	Distress state	German knowledge	Task easiness
1	,255	-,138	,055	-,361	-,370
2	**-,343**	,154	,284	**-,614**[**]	**-,458**[*]
3	-,093	-,156	,049	-,376	-,111
4	-,135	**,359**	**,393**	**-,474**[*]	**-,440**[*]

*Correlation is significant at the 0.05 level (2-tailed).
**Correlation is significant at the 0.01 level (2-tailed).

We also included the empathy related variables into this table, as some of the correlations can be considered significant (though at p<0,1): in particular empathic trait is negatively correlated with the number of mistakes in Group 2 (p=0,09) and both state-like characteristics of empathy and distress are positively correlated with the number of mistakes in Group 4 (at p=0,08 and p=0,06).

The latter result is interesting also if we look at the more general picture and compare groups with and without information about the customer. In both Groups 2 and 4 empathic state is positively correlated with the number of mistakes and for other two groups correlation is negative. Also the level of correlation between mistakes and distress is relatively higher in treatments where employees were exposed to the information about individual customer (Groups 2 and 4). In other words those employees who have information about the customer and make more mistakes in the text later report higher level of empathy and distress, while results for groups without information about the selected customer (Groups 1 and 3) do not show any relationship between these parameters.

A similar situation can be observed by looking at the relationship between empathy as trait and typing accuracy, but only for Group 2: increasing level of dispositional empathy was related to lower level of mistakes, with linear function explaining up to 28 percent (and quadratic function up to 52 percent) of variation in the amount of mistakes due to the initial differences between participants in the empathic trait.

One of the other relationships where we could observe a significant pattern of correlation between the number of mistakes and questionnaire-based variables was the parameter of social value orientation (as measured in grad)[53]. The only group where this relationship was significant was Group 1, where quadratic model explained about 33 percent of variation in the number of mistakes (Table 28).

---

[53] Parameter of Social value orientation showed a significant correlation only with the number of mistakes and not with any other main output variable.

Table 28

Number of Mistakes as a function of Social value orientation
Groups split for comparison of the linear and quadratic regression models

Group	Equation	Model Summary					Parameter Estimates		
		R Square	F	df1	df2	Sig.	Constant	b1	b2
1	Linear	,054	1,421	1	25	,244	11,497	-,124	
	Quadratic	,327	5,818	2	24	**,009**	-2,424	1,778	-,044

From Figure 12 we can clearly see that there is a significant difference between those participants who scored from 0 to 20 (calculated in grad) and those who scored between 20 and 40. This supports the division into "Individualists" (score of less than 22,45 grad) and "Prosocials" (with scores from 22, 45 to 57,15 grad).

Figure 12

Number of Mistakes as a function of Social value orientation
Graphical representation of the linear and quadratic regression model for Group 1
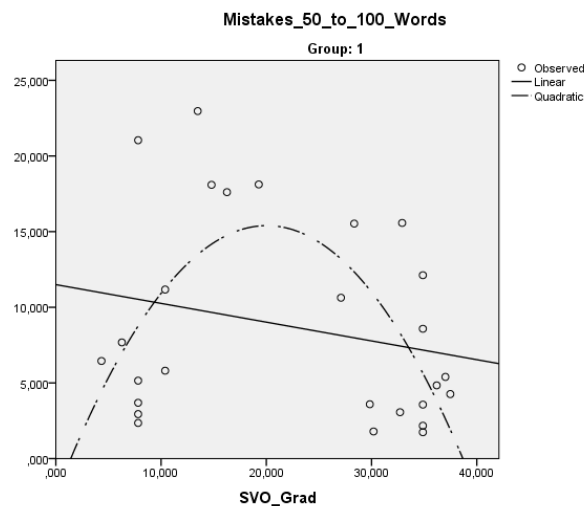
In Group 1 both "devoted" Individualists and "devoted" Prosocials made fewer mistakes than participants having middle scores. In other words being more of a "Competitive" type or closer to "Altruistic" type could be associated with stronger focus on keeping the number of mistakes in the transcript low.

Amount of Recording Transcribed[54]

Moving to the analysis of quantitative parameters of output we can see that amount of the recording transcribed was highly positively correlated with task proficiency variables in Group 1. In Group 2 it is correlated with language knowledge, but uncorrelated with perceived easiness of the task and in Group 3 the other way round. If we remember that in Group 1 participants focused on achieving the greatest preciseness, than it is not surprising that the speed of transcribing was limited by the proficiency parameters. In Group 2 the focus moved to the quantity and language knowledge became a more profound reason for variation in quantity of output. In Group 3 participants were mostly concerned about maximizing accuracy and participants perceiving task as more difficult could transcribe less.

Table 29

Groups split for correlations
of Amount of Recording Transcribed with selected questionnaire-based variables

Group	Empathic state	Distress state	Descriptive norm	German knowledge	Task easiness
1	**-,377**	**-,331**	-,018	**,575****	**,558****
2	**-,448***	-,050	**-,422***	**,477***	,150
3	,140	-,149	-,177	,339	**,539****
4	-,190	-,344	-,067	,153	,198

*Correlation is significant at the 0.05 level (2-tailed).

**Correlation is significant at the 0.01 level (2-tailed).

Also interesting is the fact that amount of the recording transcribed was significantly negatively correlated in the first two groups (especially in Group 2) with the parameter of situational empathy (but not with distress). This means that participants who felt more empathic were also the once transcribing less of the recording (or those who provided less recording reported higher level of empathic feeling). At the same time participants of Group 2 with low amount of the recording transcribed had higher scores on the question regarding helping behavior of others (descriptive norm): they believed that other people do help people with disabilities.

Interestingly situational empathy and distress also correlated in the same direction in Group 1, though at the lower significance level (at p=0,053 and p=0,092). This can be seen even clearer in the regression model with amount transcribed as dependent and descriptive norm of helping as independent variable (Table 30).

---

[54] In this part we use the parameter of amount of the recording transcribed instead of the speed of transcribing. This choice reflects the attempt to reduce the influence of scaling and rounding. Speed of transcribing being measured in the minutes of recording transcribed within one minute of the main working time, is represented by numbers less than 1; using amount of recording transcribed (multiplying the speed by 50 minutes) allows to capture more variation in the parameter.

Table 30

Amount of Recording Transcribed as a function of Descriptive norm of helping
Groups split for comparison of linear and quadratic regression models

Group	Equation	Model Summary					Parameter Estimates		
		R Square	F	df1	df2	Sig.	Constant	b1	b2
2	Linear	,403	14,828	1	22	**,001**	32,035	-4,247	
	Quadratic	,795	40,676	2	21	**,000**	72,042	-27,152	2,844

Quadratic model for Group 2 explains up to the 80 percent of variation in the amount of recording transcribed by only one variable – level of agreement with descriptive norm of helping. As we have argued above talking about preciseness the relationship between descriptive norm and these two output variables (preciseness and amount transcribed) can be interpreted in terms of participants' belief about their sole versus shared responsibility for the transcription of the recording. If other people also help, then transcribing less (but with higher quality) is fine, as "others" can complete the task. If, on the contrary, others do not help, then it is more important to transcribe as much as possible in order to provide at least some output to the customer.

Speed of Typing

If we look at the relationship between the task proficiency variables and speed of typing we will see that language knowledge was an important issue in all groups, except Group 3 (Table 31). In Group 3 as we remember there was a significant decrease in speed of typing resulting from the significant improvement of typing accuracy.

Table 31

Groups split for correlations
of the parameter of Speed of Typing with selected questionnaire-based variables

Group	Empathic trait	Empathic state	German knowledge	Task easiness	Task interest
1	,149	**-,430**[*]	**,689**[**]	**,473**[*]	-,069
2	,132	**-,480**[*]	**,691**[**]	,331	**-,408**[*]
3	**,405**[*]	,074	,339	,409	,125
4	,316	**-,431**[*]	**,695**[**]	,372	**-,412**[*]

*Correlation is significant at the 0.05 level (2-tailed).

**Correlation is significant at the 0.01 level (2-tailed).

Here we see that those who were more empathic were also the ones providing higher speed of typing. They did not necessarily make fewer mistakes, but importantly they did not dramatically reduce the speed of typing.

In all groups except Group 3 we also observe the significant correlation between speed of typing and empathic state, which we can interpret as meaning that the employees who were typing slower reported higher empathic feelings due to the participation in the experiment participants.[55]

Conclusion

Bringing together the results of the discussion in this subchapter we can conclude that participants with higher empathic abilities were providing better quality – both in terms of preciseness and the amount of mistakes - if they knew who their customer was (Groups 2 and 4).[56]

However, in Group 4 relationship between the number of mistakes and empathy is non-significant, which is also easy to explain if we remember that this is the group with employee's identity open and from the results in Group 3 we know that opening identity leads to increase in visible quality independent of other individual parameters. In other words opening the identity of the customer leads to the trade-off between speed of transcribing and preciseness and more empathic employees assign more importance to preciseness than less empathic ones. At the same time opening identity of the customer does not influence the accuracy of the transcript, but opening the identity of employee does. So, in Group 2 only more empathic employees cared about accuracy, while in Group 4 where both sides were open, both more and less empathic participants increased accuracy as the visible output parameter.

Quality in Group 2 also depends on the descriptive norm: linear function for the typing accuracy explains 11 percent and quadratic function 22 percent of the variance; and linear function for preciseness - about 40 percent. About 80 percent of the speed of transcribing in Group 2 can be explained by descriptive norm. The stronger were the norm-related beliefs, the lower was the speed of transcribing and the higher was the quality of the transcript.

5.7. Regression analysis for the main output variables

5.7.1. General overview

In this par we conduct and discuss the results of the regression analysis applied separately to each of the four main output variables. We use each output parameter (preciseness, number of mistakes, speed of transcribing and speed of typing) at the level measured after 50 minutes of the main time as the dependent variable (variable "Group_" in the regression tables). For each variable we check whether belonging to the particular treatment group has a significant effect on this variable. For that we use group belonging as independent factor variable. At the same time we control for the initial differences

---

[55] If we test this relationship by means of the regression model, the linear models with empathic state as independent variable explain about 15 percent of variation in speed of typing in Group 1 and about 18 percent in Group 2.

[56] The quadratic nature of the function can be attributed to the two extreme data points in both cases.

between participants in skills, attitudes and beliefs by using the value for the same dependent variable in the control time as a covariate. As was argued above the differences among participants in the level of output reached in the control time were the result of compound influence of preferences and situational importance of individual parameters of output as well as task proficiency of each participant. The level of output in the main time is naturally a product of the same factors, and additionally of the treatment effects. This explains why some parameters of output could change for some treatment groups, and stay the same for others.

Thus, we argue that while preferences and skills (task proficiency) are relatively stable over the time (during the experiment), it is the situational importance of different parameters of output, which is flexible and potentially altered by the interventions.[57] That is why using the value of the particular output parameters in control time as a covariate we can test in a regression model, whether the treatment information alone could be a decisive factor influencing the variability in the level of output between groups.

Further, based on our previous analysis, we check within the framework of the regression analysis, whether such parameters as dispositional empathy (trait) and adherence to the descriptive norm of helping play a significant role in explaining participants' behavior. We add these variables as covariates and analyse the resulting models. The rationale behind such narrow choice of covariates is twofold:

- First, as it was stated in the hypotheses, the difference in the direction effort in the treatments can be seen as a result of two forces – empathy and image-concerns. While the direction of effort (output variable in focus) by itself signals the difference in mechanisms behind the choice, using complimentary variables allows us to test whether individual differences in these parameters can play an additional role.
- Second, due to a relatively modest sample (24-27 participants per group), we should seek to reduce the number of variables in the model to get statistically reasonable output. That is why, without rejecting the chances that other variables could also be a part of the model, we still focus only on those parameters which have the strongest link with our hypotheses (as follows from the previous analysis).

In practical terms we fist introduce either dispositional empathy or adherence to descriptive norm (choice of parameter is based on the previous correlational analysis) as a single variable, and as a next step add an interrelation term between this parameter and a group factor reflecting belonging to the treatment group. We believe this is a necessary step, as the major argument is that different

---

[57] Of course we should take into account the change in proficiency of the participants (due to the learning effect), but remembering our focus on the direction of change as opposed to the focus on size of the output itself, we consider it constant.

information (treatment) caused participants to choose different combinations of quality and quantity and individual traits could work as either balancing or destructive force here.

All parameter estimates, as well as pairwise comparisons of the means are calculated using method of bootstrapping which allows more robust estimation in the presence of deviation from normality in the data.

5.7.2. Preciseness

We start the regression analysis from the parameter of preciseness. As we can see from Table 32 differences in the treatment information by itself do not serve as a significant predictor of the variation in preciseness in the main time (as explained above we account for the initial differences in preciseness by introducing preciseness in the control time as covariate). Individual level of the preciseness reached in the control time is a more significant predictor of participant's behavior in the main time (Model 1), than treatment group membership.[58]

Table 32

**Model 1.Tests of between-subjects effects
for regression with Preciseness as dependent variable**

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	22076,770[a]	5	4415,354	14,500	,000
Intercept	8218,073	1	8218,073	26,987	,000
Preciseness_10	18593,270	1	18593,270	61,058	,000
Speed_Words_10	1293,692	1	1293,692	4,248	,042
Group	1185,374	3	395,125	1,298	,280
Error	28624,611	94	304,517		
Total	767776,703	100			
Corrected Total	50701,382	99			

a. R Squared = ,435 (Adjusted R Squared = ,405)

However, if we look at the parameter estimates for Model 1 we can observe a clear pattern reminding of the results of non-parametric analysis of the related samples. In particular we see that, if compared

---

58 In this model we use also Speed of Typing in the control time as a covariate. The reason is that we believe that Preciseness in the control time does not reflect fully the task proficiency of the participants. As argued in the other chapters, decision about the chosen level of preciseness in the control time is based more on beliefs than on skills. Importantly preciseness is the only parameter which does not correlate with speed of typing in the control time, which allows us to put it as a covariate in the model. By including typing speed in the control time as a proxy for typing skills we want to see on the one hand whether typing skills influenced the preciseness and on the other hand, whether treatment conformation exerted significant influence on the decisions regarding preciseness if we control for the skills. We also checked the model without Speed_Words_10 as a covariate and the model had R Squared = ,410 (Adj R Squared = ,385).

to Group 1, all groups have negative B-coefficients (Table 33).[59] In other words participants from Group 1 provided the text with highest level of preciseness, which means they were the closest to the word-to-word transcripts. At the same time Group 3 provided the least precise transcripts, which as we can suggest already at this point, is a result of significant focus on the improvement of accuracy.

<div align="right">Table 33</div>

Model 1. Parameter estimates for regression with Preciseness as dependent variable

Parameter	B	Bootstrap[60]				
		Bias	Std. Error	Sig. (2-tailed)	BCa 95% CI	
					Lower	Upper
Intercept	40,313	,761	8,703	,001	22,995	59,115
Preciseness_10	,508	-,005	,083	,001	,347	,651
Speed_Words_10	,425	-,016	,204	,043	-,021	,767
Group_2	-3,527	,217	4,095	,401	-12,687	5,400
Group_3	-9,416	-,009	4,969	,067	-19,581	,378
Group_4	-6,245	,047	3,854	,125	-14,343	1,995
Group_1	0	0	0		.	.

In Model 2 we add a covariate for the agreement with the descriptive norm of helping (Table 34). In the correlation analysis we argued that level of agreement with descriptive norm of helping was positively associated with preciseness, but this relationship was significant only in Group 2.

<div align="right">Table 34</div>

Model 2.Tests of between-subjects effects
for regression with Preciseness as dependent variable

Source	Type III Sum of Squares	Df	Mean Square	F	Sig.
Corrected Model	23381,566[a]	6	3896,928	13,198	,000
Intercept	1812,174	1	1812,174	6,137	,015
Preciseness_10	18164,659	1	18164,659	61,520	,000
Speed_Words_10	1315,178	1	1315,178	4,454	,038
Descriptive norm	1413,849	1	1413,849	4,788	,031
Group	1299,408	3	433,136	1,467	,229
Error	26868,911	91	295,263		
Total	747988,696	98			
Corrected Total	50250,476	97			

a. R Squared = ,465 (Adjusted R Squared = ,430)

---

[59] Here and further we always start analysis by using Group 1 as the basis for comparison. This is full in line with the experimental design which seeks to test the differences between control treatment (Group 1) and other treatments. If later the results show that any other group differs more significantly from others, we test the model using this group as a reference group.
[60] Here and further unless otherwise noted, bootstrap results are based on 1000 bootstrap samples.

That is why in the regression analysis we first introduce this variable as a separate one, and later also add it as an interaction term with the factor of group belonging.

Model 2 has a slightly higher R Squared, than Model 1: it explains 46,5 percent of variation in preciseness (43,0 percent adjusted), while Model 1 explains 43,5 percent (40,5 percent adjusted). However, the factor reflecting the treatment group membership is still non-significant in Model 2. Parameter estimates of Model 1 (Table 35) clearly show that agreement with the descriptive norm of helping had positive influence on the preciseness (B=2,914, p<0,5).

Table 35

Model 2. Parameter estimates
for regression with Preciseness as dependent variable

Parameter	B	Bootstrap				
		Bias	Std. Error	Sig. (2-tailed)	BCa 95% CI	
					Lower	Upper
Intercept	27,404	1,513	11,547	,028	5,640	55,566
Preciseness_10	,503	-,006	,078	,001	,354	,624
Speed_Words_10	,442	-,033	,220	,043	,029	,764
Descriptive norm	2,914	-,096	1,362	,044	,341	5,317
Group_2	-3,583	,255	3,807	,362	-12,043	4,525
Group_3	-10,216	,046	4,976	,051	-20,888	,113
Group_4	-5,345	,131	4,161	,212	-13,843	2,983
Group_1	0	0	0		.	.

However, to get a clearer picture we introduce an interaction term between agreement with the descriptive norm and group membership (Model 3, Table 36). Here we can see that the parameter of group belonging becomes a significant part of the model (p<0,05), while all other variables also increase their influence.

In particular both a separate variable characterizing the agreement with descriptive norm and an interaction term are significant in this model (both at p<0,01). Also the R squared in Model 3 reaches 53 percent, which means that this model can explain about half of the variation in the dependent variable.

Table 36

Model 3.Tests of between-subjects effects
for regression with Preciseness as dependent variable

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	26645,355[a]	9	2960,595	11,037	,000
Intercept	1478,043	1	1478,043	5,510	,021
Group	3025,471	3	1008,490	3,760	,014
Preciseness_10	15990,161	1	15990,161	59,611	,000
Speed_Words_10	1518,215	1	1518,215	5,660	,020
Descriptive norm	1899,467	1	1899,467	7,081	,009
Group * Descriptive norm	3263,790	3	1087,930	4,056	,009
Error	23605,121	88	268,240		
Total	747988,696	98			
Corrected Total	50250,476	97			

a. R Squared = ,530 (Adjusted R Squared = ,482)

If we look at the parameter estimates for Model 3 (Table 37) we can conclude that treatment information influenced the level of preciseness differently for the participants with different level of agreement with the descriptive norm of helping. Also the group belonging by itself (if we account for differences in the descriptive norm) had significant influence on preciseness.

Interestingly if in Models 1 and 2 there was slightly significant difference in preciseness between Groups 1 and 3, then after introducing the interaction term (between group belonging and strength of beliefs regarding the descriptive norm) the situation changed and now we observe a significant difference (p<0,05) between the preciseness in Groups 1 and 2.

At the same time looking at the coefficients for the interaction terms itself we get a confirmation for the results of the correlation analysis: in Group 2 higher level of agreement with the descriptive norm was associated with higher preciseness. In other groups the relationship, though also positive, is less significant and in Group 4 the relationship is even negative.

If we take the interaction term into account, the information about the customer in Group 2 leads to significantly lower focus on preciseness compared to Group 1, where participants were not exposed to such information.

Table 37

**Model 3. Parameter estimates
for regression with Preciseness as dependent variable[61]**

Parameter	B	Bootstrap				
		Bias	Std. Error	Sig. (2-tailed)	BCa 95% CI	
					Lower	Upper
Intercept	45,945	1,667	11,718	,001	18,585	73,414
Preciseness_10	,476	,000	,072	,001	,331	,615
Speed_Words_10	,477	-,049	,236	,041	,032	,775
Group_1	-13,882	,090	12,699	,271	-40,475	10,947
Group_2	-43,544	,405	12,128	,001	-66,034	-16,755
Group_3	-42,668	-2,549	25,573	,097	-94,322	-,036
Group_4	0	0	0		.	.
Descriptive norm	-2,447	-,176	1,764	,164	-6,131	,476
Group_1 * Descriptive norm	4,679	-,026	2,882	,103	-,662	10,372
Group_2 * Descriptive norm	10,673	-,040	2,650	,001	5,084	15,900
Group_3 * Descriptive norm	8,716	,481	5,159	,096	-2,394	20,490
Group_4 * Descriptive norm	0	0	0		.	.

From the Table 37 we can finally conclude that among all groups Group 4 was the least influenced by the agreement with the descriptive norm. At the same time Group 2 was the most influenced by this parameter: those participants who believed that other people also help those with disabilities had much higher preciseness than those who did not believe in this statement. If we account for the descriptive norm, than we see that Group 4 had the highest preciseness (though Group 1 was not significantly different from it) and Group 2 had a significantly lower preciseness than Group 4.

If we look at the means for the groups calculated using the parameters of Model 3 (Table 38) we can conclude that Group 1 had the highest mean preciseness and also the tightest (BCa) confidence interval among all groups. Groups 2 and 4 were close in the preciseness to each other and also relatively close to Group 1. Group 3 was lower in preciseness than each of the other groups and also had the widest confidence interval.

---

[61] Since the analysis of Model 3 shows that in fact Group 4 (and not Group 1) has the highest level of preciseness (if we account for the interaction between differences in the agreement with descriptive norm and group belonging), we use Group 4 as the reference group for parameter estimates and further analysis (Table 37).

Table 38

Model 3. Estimates of means
for regression with Preciseness as dependent variable

Group	Mean	Std. Error	95% CI		Bootstrap for Mean			
			Lower Bound	Upper Bound	Bias	Std. Error	BCa 95% CI	
							Lower	Upper
1	89,130[a]	3,172	82,826	95,433	-,060	2,441	83,828	93,667
2	85,704[a]	3,351	79,044	92,364	,082	3,294	78,490	92,136
3	78,015[a]	3,498	71,063	84,966	-,456	5,574	65,790	87,361
4	82,528[a]	3,375	75,821	89,235	,048	3,453	75,047	89,204

a. Covariates appearing in the model are evaluated at the following values: Preciseness_10 = 80,443, Speed_Words_10 = 18,910, Descriptive norm = 4,38.

In order to check whether the differences in preciseness among groups are significant, we perform pairwise comparisons of the means calculated in Model 3 (Table 39). All groups have a positive difference between their mean preciseness and the one for Group 1, which in this case means that preciseness in Group 1 is indeed the highest. However, this difference reaches slight significance only in Group 3 (p<0,1), which leads us to the conclusion that the requirement of non-anonymity of the employee moves focus away from preciseness towards some other parameters of output.

Table 39

Model 3. Pairwise comparisons
for regression with Preciseness as dependent variable

(I) Group	(J) Group	Mean Difference (I-J)	Bootstrap				
			Bias	Std. Error	Sig. (2-tailed)	BCa 95% CI	
						Lower	Upper
1	2	3,426	-,143	3,601	,351	-3,287	9,915
	3	11,115	,396	5,720	,063	1,028	24,246
	4	6,602	-,109	4,018	,111	-1,387	14,151

Interestingly, according to Model 3, participants of Group 2 provide the transcripts not significantly less precise than those of Group 1. This means that in Group 2 participants' beliefs about the helping behavior of others (descriptive norm of helping) were decisive for the level of preciseness they provided: those participants who believed that others also help invested more time and effort to provide more precise transcripts. As a result they could reach such level of preciseness that compensated for the lack of preciseness of those participants who did not believe that others help people with disabilities.

If we use the same parameters as in Model 3 for testing the differences between changes in preciseness from control to the main time (instead of absolute level of preciseness reached in the main time), we can see that Group 1 had the highest improvement in preciseness (on average almost 9 words per minute of recording) and it was the only group with strictly positive confidence interval, while Group 3 had high chances to decrease the preciseness (on average about 2,5 words per minute of recording) (Table 40).[62]

<div align="right">Table 40</div>

Model 3. Estimates of means
for regression with change in Preciseness as dependent variable

Group	Mean	Std. Error	95% CI		Bootstrap for Mean			
			Lower Bound	Upper Bound	Bias	Std. Error	BCa 95% CI	
							Lower	Upper
1	8,676a	3,173	2,371	14,981	-,154	2,914	2,908	13,863
2	5,264[a]	3,352	-1,397	11,925	,126	3,426	-1,475	12,096
3	-2,432[a]	3,499	-9,385	4,521	-,506	5,254	-13,017	5,980
4	2,083[a]	3,376	-4,626	8,791	-,085	3,671	-6,151	9,238

a. Covariates appearing in the model are evaluated at the following values: Preciseness_10 = 80,443, Speed_Words_10 = 18,910, Descriptive norm = 4,38.

This analysis brings us to the following conclusions:

1. Preciseness in the main time was strongly related to the level of agreement with the descriptive norm of helping. Generally the stronger the participants believed that others help people with disabilities, the higher was the preciseness of their transcripts. The relationship was driven by the strong relationship in Group 2, while it was the weakest (to the level of being non-existent) in Group 4.

2. If we account for the influence of interaction between descriptive norm and group membership, than Group 4 had the highest chances of providing the most precise transcripts and Group 2 the least precise.

3. If we account for the influence of the interaction between group belonging and descriptive norm, then participants of Group 1 provided the most precise transcripts of all groups and Group 3 the least precise.

Thus, preciseness being on one hand the parameter prescribed by the instructions and on the other not easily visible and arguably important to maximize, depended on the treatment information. However, agreement with descriptive norm was very important for the anonymous participants with the

---

[62] Further test shows, that the difference between Groups 1 and 3 is significant (p=0,064).

information about customer. As a result participants with general information provided the best preciseness as they were paying attention to the requirements. In Group 2 employees split according to their beliefs about descriptive norm. In Group 3 participants also split, but the reason was not the norm (what was the reason is the topic of further discussion). In Group 4 agreement with the norm did not have a significant influence on preciseness (or we can talk even about negative influence).

5.7.3. Number of Mistakes

In the regressions below we use the number of mistakes (per 100 words of the transcript) as the dependent variable and show that, in agreement with the results of the previous analysis, Group 3 made the fewest mistakes among all groups, while Groups 1 and 2 made the most mistakes.

We first start with the simple model looking for the differences in accuracy related to the difference in the treatment information provided to the groups (Table 41). We find that belonging to the treatment group was not a significant predictor of accuracy in Model 1, while much of the variability can be explained by the number of mistakes made by the participant in the control time (hence the high level of the R Squared – about 70 percent - for this model).

Table 41

Model 1. Tests of between-subjects effects
for regression with Number of Mistakes as dependent variable

Source	Type III Sum of Squares	Df	Mean Square	F	Sig.	Partial Eta
Corrected Model	3071,546[a]	4	767,887	53,333	,000	,692
Intercept	61,160	1	61,160	4,248	,042	,043
Group	84,931	3	28,310	1,966	,124	,058
Mistakes_10	2998,969	1	2998,969	208,291	,000	,687
Error	1367,805	95	14,398			
Total	10048,867	100				
Corrected Total	4439,351	99				

a. R Squared = ,692 (Adjusted R Squared = ,679)

However, , if we look at the parameter estimates for Model 1 (Table 42) we can see that compared to Group 1 every other group has a lower number of mistakes. However, only in Group 3 this difference is significant (p<0,05).

Table 42

**Model 1. Parameter estimates
for regression with the Number of Mistakes as dependent variable**

Parameter	B	Bias	Std. Error	Sig. (2-tailed)	BCa 95% CI Lower	Upper
					Bootstrap	
Intercept	1,929	-,021	,815	,025	,419	3,512
Group_2	-,067	,048	1,147	,956	-2,264	2,354
Group_3	-2,302	,026	1,025	,027	-4,315	-,301
Group_4	-,592	-,039	1,015	,562	-2,507	1,174
Group_1	0	0	0			
Mistakes_10	,672	,001	,061	,001	,547	,798

Further, as with the analysis of preciseness, we add several individual characteristics (questionnaire-based) as covariates. In the previous analysis of the relationships between output variables and parameters reflecting individual differences between participants it was shown that number of mistakes was related to the empathy as trait and SVO. However, group-level analysis revealed that these relationships were significant only for some groups and link was not particularly strong. In the present analysis inclusion of these variables did not render significant improvement of the model. In particular, the empathic trait was not a significant in the model ($p > 0,1$) and its addition did not lead to any improvement of the model (R squared=0,693, Adjusted R Squared=0,676).

In order to keep the regression models provided here more straightforward (also due to the specifics of dataset discussed elsewhere) we decided not to include these variables in the regression analysis and proceeded with a simple Model 1.

Further we move to the comparisons of mean values calculated using the values of parameters in the experimental sample. Table 43 shows that Group 1 had the highest average amount of mistakes (based on the estimates of Model 1), but it was almost the same as in Group 2. Groups with the requirement to open employee's identity (Group 3 and 4) had lower amount of mistakes and this is especially true for Group 3, which apparently had the lowest number of mistakes among all groups. At the same time, as in Model 1 for preciseness, Group 2 had the largest confidence interval for the estimates of the group mean for the amount of mistakes.

Table 43

**Model 1. Estimates of means
for regression with Number of Mistakes as dependent variable**

Group	Mean	Std. Error	Bootstrap for Mean			
			Bias	Std. Error	BCa 95% CI	
					Lower	Upper
1	8,201[a]	0,731	,003	,873	6,493	9,898
2	8,134[a]	0,762	,052	1,085	6,113	10,431
3	5,899[a]	0,775	,029	,792	4,223	7,620
4	7,609[a]	0,775	-,036	,956	5,860	9,394

a.Covariates appearing in the model are evaluated at the following values: Mistakes_10_to_100_Words = 9,33327.

The pairwise comparisons (Table 44) support the conclusion regarding the link between non-anonymity of the employee and focus on accuracy: both Group 1 and Group 2 have significantly higher number of mistakes than Group 3 (p<0,05), while the difference between Groups 3 and 4 is insignificant (p>0.1).

Table 44

**Model 1. Pairwise comparisons
for regression with Number of Mistakes as dependent variable**

(I) Group	(J) Group	Mean Difference (I-J)	Bootstrap				
			Bias	Std. Error	Sig. (2-tailed)	BCa 95% CI	
						Lower	Upper
3	2	-2,236	-,022	1,147	,047	-4,477	-,102
	4	-1,711	,065	1,037	,113	-4,037	,390
	1	-2,302	,026	1,025	,027	-4,315	-,301

If we use the same parameters as in Model 1 for testing the differences between changes in the amount of mistakes from control to the main time (instead of absolute number of mistakes in the main time), we can see that Group 3 has the highest reduction in the amount of mistakes (on average 2,4 mistakes per 100 typed words less) (Table 45). Importantly Group 4 also has a large mean value for the change in the amount of mistakes (at least larger than for Groups 1 and 2) and both Group 3 and Group 4 are the groups with strictly negative confidence interval, which means that they definitely improved accuracy of their transcripts, while Groups 1and 2 have intervals including zero.

Table 45

Model 1. Estimates of means
for regression with change in Number of Mistakes as dependent variable

| Group | Mean | Std. Error | 95% CI | | Bootstrap for Mean | | | |
| | | | Lower Bound | Upper Bound | Bias | Std. Error | BCa 95% CI | |
							Lower	Upper
1	-,924[a]	,584	-2,083	,235	,002	,595	-2,130	,238
2	-,922[a]	,620	-2,153	,309	-,002	,770	-2,415	,667
3	-2,348[a]	,632	-3,603	-1,094	,050	,543	-3,487	-1,180
4	-1,388[a]	,618	-2,616	-,161	,003	,629	-2,623	-,054

a. Covariates appearing in the model are evaluated at the following values: Mistakes_10_to_100_Words = 9,25404.

Further tests showed that the difference between Groups 1 and 3 was significant ($p<0,1$) supporting the previous observation regarding the strong unified focus of Group 3 on improving the accuracy of the transcripts (in terms of mistakes).

Thus, summing up the results for this sub-chapter we can conclude that:

- Participants of Group 3 provided the transcripts with the fewest mistakes among all groups;
- Parameters characterizing individual differences (in particular, empathy) did not play a significant role in explaining differences between treatments in the amount of mistakes.
- Groups with non-anonymous participants strictly decreased the amount of mistakes, while anonymous participants were less united and did not achieve the significance threshold.

The results of the regression analysis for both parameters characterizing the quality of output (preciseness and accuracy of the transcript) allow for the following conclusions:

1. Participants having only general information about the customer and working anonymously (Group 1) were the best in terms of preciseness and one of the worse in terms of accuracy. Thus, they maximized the level of parameter required by the instructions (preciseness) and did not focus on accuracy.
2. For participants working anonymously but having the information about the selected customer (Group 2) level of preciseness significantly positively depended on the agreement with the descriptive norm of helping. In fact, though, they were not significantly different in preciseness from Group 1 (best) and significantly worse in accuracy (number of mistakes) than Group 3 (best).
3. Participants having only general information about the customer, but required to open their identity (Group 3) were the worst in terms of preciseness and the best in terms of accuracy. This supports our view that the requirement of non-anonymity moved focus from maximizing the required parameter of quality to highly visible parameter of quality.

4. Participants having the information about individual customer in Groups 2 and 4 were providing the transcripts of similar accuracy and preciseness. However, participants working anonymously (Group 2) were slightly better in preciseness and non-anonymous participants (Group 4) – in accuracy.

5.7.4. Amount of Recording Transcribed

Further we analyse the regressions for two parameters characterizing the quantity of output. The first one is the amount of the recording transcribed within the main time.[63]

As in the previous analysis, in this regression we use the factor of belonging to the treatment group by itself as an independent variable, and further introduce also an interaction term with the covariate of descriptive norm.

Table 46

Model 1.Tests of between-subjects effects
for regression with Amount of Recording Transcribed as dependent variable

Source	Type III Sum of Squares	Df	Mean Square	F	Sig.
Corrected Model	3936,228[a]	4	984,057	22,185	,000
Intercept	686,784	1	686,784	15,483	,000
Speed_Minutes_10	3612,701	1	3612,701	81,445	,000
Group	129,141	3	43,047	,970	,410
Error	4213,958	95	44,357		
Total	24766,943	100			
Corrected Total	8150,187	99			

a. R Squared = ,483 (Adjusted R Squared = ,461)

As we can see from the Table 46, the factor of belonging to the treatment group is not significant in this model. However, due to the inclusion of the parameter of transcribing speed in the control time we come up with the model having relatively high R Squared of more than 45 percent.

B-coefficient for the speed of transcribing in the control time is positive, meaning that higher initial speed is associated with higher amount transcribed: in other words the faster the participant was in the control time, the more recording she could transcribe in the main time. Parameter estimates for the group membership show that anonymous participants having the information about selected customer

---

[63] We use this parameter instead of speed of transcribing, as it easier for interpretation and it has more appropriate scale.

(Group 2) transcribed more recording (in minutes), than participants of the other groups (p=0,073 and strictly positive confidence interval) (Table 47).

Model 1. Parameter estimates
for regression with Amount of Recording Transcribed as dependent variable

Parameter	B	Bootstrap				
		Bias	Std. Error	Sig. (2-tailed)	BCa 95% CI	
					Lower	Upper
Intercept	4,052	-,038	2,045	,084	,719	7,883
Speed_Minutes_10	28,662	,257	8,300	,003	11,563	43,693
Group_2	2,420	-,128	1,245	,073	,287	4,433
Group_3	-,038	,040	1,547	,981	-3,161	3,331
Group_4	-,522	,001	1,375	,726	-3,157	2,092
Group_1	0	0	0		.	.

Following the pattern of the analysis for the preciseness and our conclusion made in the section__ regarding the relationship between amount of recording transcribed and agreement with the descriptive norm, we construct a new model – Model 2 –where we add first the covariate of descriptive norm (Table 48) and later (Model 3) also the interaction term between treatment group membership and descriptive norm (beliefs regarding others helping behavior) to the regression model.

Model 2. Tests of between-subjects effects
for regression with Amount of Recording Transcribed as dependent variable

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	4184,346[a]	5	836,869	19,618	,000
Intercept	445,049	1	445,049	10,433	,002
Speed_Minutes_10	3381,480	1	3381,480	79,269	,000
Descriptive norm	117,909	1	117,909	2,764	,100
Group	129,211	3	43,070	1,010	,392
Error	3924,564	92	42,658		
Total	24602,081	98			
Corrected Total	8108,910	97			

a. R Squared = ,516 (Adjusted R Squared = ,490)

As we can see the introduction of the descriptive norm into the model improves the R squared which means that a new model is able to explain about 50 percent of the variation in the dependent variable – amount of the recording transcribed. Adding the interaction term (as explained above) does not improve the model (R Squared = ,538; Adjusted R Squared = ,497), but it is interesting to see, that if we account for the difference in the interaction effect between group membership and agreement with descriptive norm, Group 2 provided significantly higher amount of recording transcribed, than participants of Group 4 (Table 49). At the same time the norm itself played more important role in Group 2 than in Group 4 ($p<0,1$): participants with higher level of agreement with descriptive norm were transcribing less recording, than participants with lower level of such agreement.

Table 49

Model 3. Tests of between-subjects effects
for regression with Amount of Recording Transcribed as dependent variable[64]

Parameter	B	Bootstrap				
		Bias	Std. Error	Sig. (2-tailed)	BCa 95% CI	
					Lower	Upper
Intercept	3,657	-,171	3,020	,260	-2,228	9,373
Group_2	13,929	-1,358	5,964	,040	3,403	21,445
Group_3	3,049	1,897	11,740	,817	-25,126	30,172
Group_4	2,106	-,033	3,725	,580	-4,670	9,341
Group_1	0	0	0		.	.
Speed_Minutes_10	27,167	,456	8,261	,006	10,052	43,044
Descriptive norm	,167	,013	,495	,752	-,775	1,131
Group_2 * Descriptive norm	-2,629	,261	1,231	,063	-5,378	,564
Group_3 * Descriptive norm	-,509	-,396	2,464	,852	-4,434	2,836
Group_4 * Descriptive norm	-,598	,022	,870	,508	-2,478	1,100
Group_1 * Descriptive norm	0	0	0		.	.

Estimates of the means based on Model 3 (Table 50) show that Group 2 is characterized by the highest average amount of minutes transcribed (about 15 minutes of recording were transcribed on average by the participants of this group in the 50 minutes of the main time). Group 4 had the lowest average amount transcribed (about 11 minutes of recording transcribed). However, none of these differences was significant.

---

[64] Model 3 is similar to Model 2, but includes also the interaction term between the factor of group membership and agreement with the descriptive norm.

Table 50

**Model 3. Estimates of the means
for regression with Amount of Recording Transcribed as dependent variable**

Group	Mean	Std. Error	95% CI		Bootstrap for Mean			
			Lower Bound	Upper Bound	Bias	Std. Error	BCa 95% CI	
							Lower	Upper
1	12,270[a]	1,265	9,757	14,784	-,009	,914	10,439	14,140
2	14,691[a]	1,333	12,042	17,339	-,194	1,328	12,365	16,739
3	13,091[a]	1,417	10,276	15,907	,144	1,863	9,126	17,043
4	11,760[a]	1,336	9,105	14,415	,038	1,407	9,392	14,384

a. Covariates appearing in the model are evaluated at the following values: Speed_Minutes_10 = ,29020408, Descriptive norm = 4,38.

If we use the same parameters as in Model 3 for testing the differences between changes in the speed of transcribing from control to the main time (instead of absolute amount of the recording transcribed in the main time) (Table 51), we can see that Groups 1 and 4 have strictly negative confidence intervals which confirm that these groups reduced the speed of transcribing in the main time (on average for 0,5 minutes of recording in 10 minutes of working time).

Table 51

**Model 3. Estimates of means for regression
with change in Speed of Transcribing as dependent variable**

Group	Mean	Std. Error	95% CI		Bootstrap for Mean			
			Lower Bound	Upper Bound	Bias	Std. Error	BCa 95% CI	
							Lower	Upper
1	-,045[a]	,025	-,095	,005	,000	,019	-,085	-,008
2	,004[a]	,027	-,049	,057	-,004	,023	-,041	,037
3	-,028[a]	,028	-,085	,028	,003	,036	-,121	,046
4	-,055[a]	,027	-,108	-,002	,000	,029	-,110	-,001

a. Covariates appearing in the model are evaluated at the following values: Speed_Minutes_10 = ,290, Descriptive norm = 4,38.

The above analysis brings us to the conclusion that the participants of Group 2 provided the highest amount of the recording transcribed of all groups. However, beliefs about others' helping behaviour played an important role in this group: those participants who believed that others also help people with disabilities transcribed less than those who did not. Combining this result with the previous discussion of the parameter of preciseness we come to the conclusion that those participants who were exposed to the information about individual customer (with disabilities), but believed that others did not help people with disabilities, were more focused on maximizing the amount transcribed and were

ready to sacrifice preciseness for it. Those who had positive beliefs about others' behaviour transcribed less recording, but made their transcripts more precise.

5.7.5. Number of Words Typed[65]

Analysis of the regressions for the number of words typed also confirms our previous conclusions regarding the differences in situational importance and focus among groups. In particular we show that Group 2 typed the highest number of words, but the difference between Groups 2 and 1 is negligible if we take account of the differences in the individual levels of the empathic trait.

As we argued above, speed of typing (or number of words typed) was the most general and clear indicator of the effort exerted by each participant. This is because it is connected, closer than any other output variable, to the task proficiency of the participant: seeing proficiency in typing and language knowledge as relatively stable within the experimental session we believe that higher speed of typing (if we take account of the speed of typing in the control time) can be associated with higher effort. Even if we assume a short-term learning effect, which leads to increase in speed of typing, we interpret it also as a sign of increase in motivation (both learning and applying the knowledge requires extra effort).

Indeed, if for other output variables belonging to a particular treatment group was not by itself significantly improving the simple regression model explaining the variable of interest, for amount typed it does (Table 52).

Table 52

Model 1.Tests of between-subjects effects
for regression with Number of Words Typed as dependent variable

Source	Type III Sum of Squares	Df	Mean Square	F	Sig.
Corrected Model	10743038,657[a]	4	2685759,664	40,340	,000
Intercept	2564287,434	1	2564287,434	38,516	,000
Group	812568,059	3	270856,020	4,068	,009
Speed_Words_10	10479929,251	1	10479929,251	157,410	,000
Error	6324853,053	95	66577,401		
Total	144696879,000	100			
Corrected Total	17067891,710	99			

a. R Squared = ,629 (Adjusted R Squared = ,614)

---

[65] As in the section devoted to the analysis of the amount of the recording transcribed in this section we use the number of words typed within the main time instead of individual speed of typing. The arguments brought in the footnotes regarding the amount of the recording transcribed generally apply.

In Model 1 if we account for the differences in initial speed of typing, differences in treatment information were a significant predictor of amount typed within the main time. From the Table 53 we see that compared to Group 1 only Group 3 had significantly lower amount transcribed, while Group 4 had slightly lower and Group 2 slightly higher amount typed than Group 1.

Model 1. Parameter estimates
for regression with Number of Words Typed as dependent variable

Parameter	B	Bootstrap				
		Bias	Std. Error	Sig. (2-tailed)	BCa 95% CI	
					Lower	Upper
Intercept	455,956	-8,735	103,477	,001	247,871	639,727
Group_2	39,888	-1,926	62,006	,534	-77,849	156,807
Group_3	-200,264	1,775	77,779	,015	-363,479	-43,701
Group_4	-79,035	,845	64,585	,231	-192,778	35,904
Group_1	0	0	0		.	.
Speed_Words_10	38,240	,524	5,106	,001	29,191	49,387

Following the conclusion of the previous analysis we introduce some individual characteristics (questionnaire-based) into the regression model and test them for significance. First variable of interest is Empathic trait, which presence, as we see from the Table 54, significantly improves the model.

Model 2.Tests of between-subjects effects
for regression with Number of Words Typed as dependent variable

Source	Type III Sum of Squares	Df	Mean Square	F	Sig.
Corrected Model	11539454,669[a]	5	2307890,934	39,241	,000
Intercept	18974,169	1	18974,169	,323	,571
Group	683491,447	3	227830,482	3,874	,012
Speed_Words_10	10891879,964	1	10891879,964	185,195	,000
Empathic trait	796416,011	1	796416,011	13,541	,000
Error	5528437,041	94	58813,160		
Total	144696879,000	100			
Corrected Total	17067891,710	99			

a. R Squared = ,676 (Adjusted R Squared = ,659)

Empathic trait, as we see from the value of the B-coefficient, had a significant positive relationship with the number of words typed (Table 55). This means that more empathic participants (those scoring higher on the empathic trait questionnaire) typed more words, which in the presence of control

variable for task proficiency (we use speed of typing in the control time as a proxy for task proficiency) can be interpreted in terms of higher effort. It is interesting that introduction of this parameter led to the change in the sign of the B-coefficient for Group 2. In other words if we account for empathic trait, all groups were providing lower amount of transcript than Group 1.

<div align="right">Table 55</div>

Model 2. Parameter estimates
for regression with Number of Words Typed as dependent variable

Parameter	B	Bootstrap				
		Bias	Std. Error	Sig. (2-tailed)	BCa 95% CI	
					Lower	Upper
Intercept	11,750	-1,262	134,501	,937	-254,645	269,899
Group_2	-33,676	,636	67,744	,617	-165,110	102,686
Group_3	-214,261	6,680	74,497	,007	-374,503	-44,671
Group_4	-123,724	1,530	66,534	,074	-255,323	23,307
Group_1	0	0	0		.	.
Speed_Words_10	39,119	,709	4,065	,001	31,777	49,336
Empathic trait	37,540	-1,080	12,350	,003	13,879	58,850

In order to understand the relationship between the empathic trait and amount typed we also modify Model 2 and add an interaction term between the factor of treatment group membership and empathic trait (Table 56).

<div align="right">Table 56</div>

Model 3. Tests of between-subjects effects
for regression with Number of Words Typed as dependent variable

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	11937948,887[a]	8	1492243,611	26,471	,000
Intercept	1259,996	1	1259,996	,022	,881
Group	628609,755	3	209536,585	3,717	,014
Speed_Words_10	11287661,751	1	11287661,751	200,232	,000
Empathic trait	407467,613	1	407467,613	7,228	,009
Group * Empathic trait	398494,219	3	132831,406	2,356	,077
Error	5129942,823	91	56372,998		
Total	144696879,000	100			
Corrected Total	17067891,710	99			

a. R Squared = ,699 (Adjusted R Squared = ,673)

In the Table 57 we can see that the only group which has significant difference with Group 1 is still Group 3. In other words if participants were exposed to the requirement of non-anonymity (but did not

have the information about customer) they were generally typing less than their anonymous counterparts. However, among such non-anonymous employees there were significant differences in the number of words typed related to their empathic trait: those more empathic typed more (the interaction term for the Group 3 is positive and significant at p<0.05).

Model 3. Parameter estimates
for regression with Number of Words Typed as dependent variable

Parameter	B	Bootstrap				
		Bias	Std. Error	Sig. (2-tailed)	BCa 95% CI	
					Lower	Upper
Intercept	376,767	-28,815	232,414	,099	-66,566	724,526
Group_2	-291,811	10,400	342,560	,371	-977,392	436,787
Group_3	-942,292	44,265	342,892	,008	-1699,700	-105,580
Group_4	-366,863	-27,252	444,259	,389	-1285,873	431,478
Group_1	0	0	0		.	.
Speed_Words_10	40,569	,538	3,534	,001	33,878	49,172
Empathic trait	3,235	1,704	19,952	,878	-30,967	51,741
Group_2 * Empathic trait	24,303	-1,079	28,641	,382	-30,177	71,834
Group_3 * Empathic trait	62,682	-3,633	27,117	,024	13,314	104,212
Group_4 * Empathic trait	22,457	1,944	36,932	,528	-48,726	107,113
Group_1 * Empathic trait	0	0	0		.	.

A closer look at the estimated means for the amount typed in Model 3 reveals that the mean values for Groups 1 and 2 are very close to each other (Table 58). Group 3 is characterized by the lowest value of the mean with confidence interval values located relatively lower than those for other groups. Also if we look at the confidence intervals we can observe a difference in the their levels for groups with and without the requirement of non-anonymity of employee: participants working non-anonymously had the means in the (BCa) interval of 894-1195 words (typed in 50 minutes of the main time) and anonymous participants had the mean in the much lower interval of 1066-1324 words (typed in 50 minutes of the main time).

In the Table 59 we confirm that the difference between means for the number of words typed are significant for Groups 3 and 1 and Groups 3 and 2.

Table 58

**Model 3. Estimates of means
for regression with Number of Words Typed as dependent variable**

| Group | Mean | Std. Error | 95% CI | | Bootstrap for Mean | | | |
| | | | Lower Bound | Upper Bound | Bias | Std. Error | BCa 95% CI | |
							Lower	Upper
1	1191,752[a]	48,396	1095,619	1287,886	,051	62,914	1065,850	1324,078
2	1197,648[a]	57,031	1084,364	1310,933	-1,953	50,190	1110,672	1281,346
3	1017,319[a]	49,326	919,339	1115,300	1,695	61,256	894,097	1137,309
4	1099,985[a]	49,260	1002,136	1197,834	-2,422	53,966	1010,268	1195,015

a. Covariates appearing in the model are evaluated at the following values: Speed_Words_10 = 19,112, Empathic trait = 12,250.

From the Table 59 we can conclude that both Groups 1 and 2 were providing significantly longer transcripts than Group 3. There is also an indication that Group 4 was slower than Group 2 (but not Group 1, which supports the view that Group 2 indeed transcribed more than Group 1).

Table 59

**Model 3. Pairwise comparisons
for regression with Number of Words Typed as dependent variable[66]**

| (I) Group | (J) Group | Mean Difference (I-J) | Bootstrap | | | | |
| | | | Bias | Std. Error | Sig. (2-tailed) | BCa 95% CI | |
						Lower	Upper
	2	-180,329	,393	64,257	,009	-315,746	-60,695
3	4	-82,666	4,587	67,957	,230	-221,637	68,782
	1	-174,433	1,890	71,871	,022	-334,911	-26,504

If we use the same parameters as in Model 3 for testing the differences between changes in the speed of typing from control to the main time (instead of absolute amount of words typed in the main time), we can see that Group 3 has the highest reduction in the speed of typing (on average they typed 2,6 words per minutes less) (Table 60). Importantly this is the only group with confidence interval not including zero, which means that we can be sure that participants of this group indeed reduced the speed of typing.

---

[66] Also the univariate F test based on the linearly independent pairwise comparisons among the estimated marginal means shows the significant difference in the amount of words typed among groups (F=2,838, p=0,042).

Table 60

Model 3. Estimates of means
for the regression with change in Number of Words Typed as dependent variable

Group	Mean	Std. Error	Bootstrap for Mean			
			Bias	Std. Error	BCa 95% CI	
					Lower	Upper
1	,897[a]	,968	,013	,997	-1,189	2,903
2	1,023[a]	1,141	,025	,733	-,542	2,666
3	-2,586[a]	,987	-,020	1,195	-5,127	-,220
4	-,938[a]	,986	-,025	,898	-2,661	,759

a. Covariates appearing in the model are evaluated at the following values: Speed_Words_10 = 19,112, Empathic trait = 12,250

The pairwise comparisons (Table 61) confirm the conclusion that Groups 1 and 2 had significantly higher change in the speed of typing than Group 3.

Table 61

Model 3. Pairwise comparisons
for regression with change in Speed of Typing as dependent variable

(I) Group	(J) Group	Mean Difference (I-J)	Bootstrap				
			Bias	Std. Error	Sig. (2-tailed)	BCa 95% CI	
						Lower	Upper
3	2	-3,610	-,045	1,274	,008	-6,278	-1,218
	4	-1,648	,005	1,347	,226	-4,317	,902
	1	-3,483	-,033	1,484	,026	-6,516	-,687

This discussion brings us to the following conclusions:

- Participants in the groups without the requirement to open their identity (Groups 1 and 2) typed significantly more words. Interestingly, while Group 2 provided the highest amount of transcript in terms of minutes of recoding transcribed, its leadership in terms of the number of words typed is not as clear. In fact Group 1 was almost as fast as Group 2 and typed comparable number of words. This can mean that due to the lower preciseness Group 2 could achieve higher speed of transcribing.

- Group 3 had the lowest number of words typed, which in the light of previous discussion can be explained by the focus on visible parameters of quality (minimizing the amount of mistakes), which influenced the typing speed of participants. At the same time those participants in Group 3 who were scoring higher on the parameter of empathic trait also provided longer transcripts.

5.8. Analysis of the behaviour in the additional time

5.8.1. General description of the procedure

After main time was over the participants were offered an opportunity to stay for additional 30 minutes and continue working on the task. This time was not paid. Participants were provided with the last set of instructions where the special focus was directed towards the importance of quality for the (potential) customer (Figure 13). Quality in this context has been explained in terms of typing accuracy (low number of mistakes) as well as proper formatting (paragraphing, numbering where necessary, etc.)

Figure 13

Selected parts from the Instructions provided in the additional time[67]
(part identical for all groups)

> … In case you decide to continue, please keep in mind that not only the quantity of transcribed recording, but also the quality of the text which is the output of your effort is of importance for beneficiary. In particular, the text is readable if
>
> - There are as few as possible typos
> - There is a clear structure introduced by paragraphs and if necessary numbering

We see the main purpose of this part of experimental procedure in its potential to answer the following questions:

- What are the chances that participants in our random experimental sample will be willing to work without any financial reward? Are there are differences among treatment groups in the probability of exerting such unpaid effort (taking decision to stay and work in the additional time) and its intensity?
- Which parameters of output come into focus? Are there any differences among treatment groups in such focus?

Thus, this part was designed in such a way that participants had to take three decisions:

1. To stay or to leave?

2. If stayed, then working hard or hardly working?

2. While working, focus on improving output quantity or quality?

Thus, this experimental part was aiming at clarifying the results and the interpretations of the main part.

---

[67] Full text of the instructions for the additional time is provided in the Appendix 6.

In particular we argued that compared to the control group (Group 1):

- The exposure to the information about an individual customer (Group 2), leads to a higher willingness to stay in the unpaid period and provision of higher effort.

- The requirement of opening employee's identity (in the absence of the information about an individual customer) increases willingness to stay in the unpaid period, but not necessarily results in the higher overall level of effort (due to the desire to "show" the presence and the effort, but not to benefit the customer);

- Combined exposure to the treatment effects results in the highest chances of staying in the unpaid time with high general level of effort.

5.8.2. Analysis of the data for the decision to stay

More than 70 percent of participants (73 out of 100) stayed in the additional unpaid time for the period from a few minutes to the full allowed duration of 30 minutes.[68] In other words almost three fourth of the sample used the opportunity to "improve" their transcripts without being paid for that.

As we see from the Table 62, there is a clear difference in proportion of those who stayed and not stayed between the groups: while in Group 2 only slightly more than 50 percent of the participants stayed, in Group 4 this ratio is close to 90 percent. Generally even from this analysis it is clear that those who were in the groups with the requirement to open employee's identity had higher probability of staying than those who worked anonymously.[69]

Table 62

**Groups split for the Number of participants who stayed versus did not stay
in the additional time**

Group	Stayed	Did not stay	Total
1	19	8	27
2	14	11	25
3	19	5	24
4	21	3	24
Total	73	27	100

---

[68] We consider as "stayed" all those who made any changes to their transcript after main time, i.e. those for whom there is a difference (e.g. in the number of words typed; in formatting, etc.) between the text saved as output at the end of main time and the final transcript.

[69] We admit that there is a number of other factors which could influence the decisions of participants regarding leaving or staying in the additional time. In particular, observing the behavior of others in the group individual participants could take the decision to leave or stay (group pressure effect). The effect of such factors should be either controlled for or explicitly tested in the future studies.

If we combine the groups with the information about employee required and those without such requirement (Figure 14), we can see that for the first ones the proportion of those who stayed to those who did not stay is 40 to 8, while for the second ones it is 33 to 20.

Number of participants who stayed / did not stay in the additional time
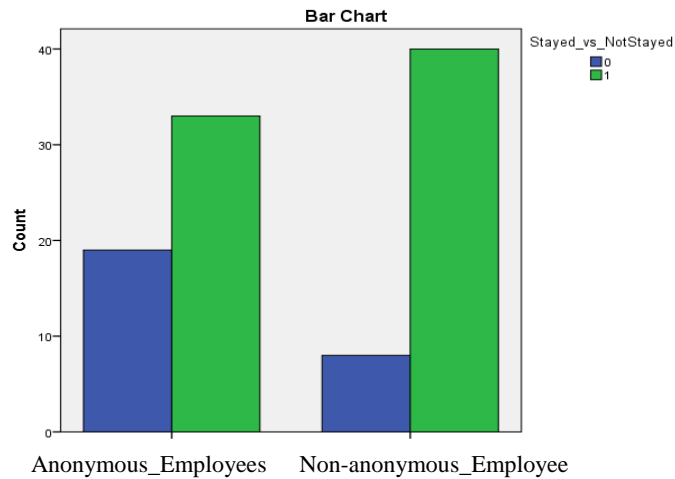
In order to have statistical support for this difference we conducted the logistic regression with dichotomous dependent variable "Stayed". Table 63 shows that participants in Group 2 had significantly lower chances to stay in the optional time, than the participants of Group 4. As the only difference between these groups was in non-anonymity of the employee we can attribute this behaviour to this treatment effect.

Model 1. Parameter estimates
for logistic regression with binary decision variable "Stayed" as dependent variable

	B	S.E.	Wald	Df	Sig.	Exp(B)
Group			6,299	3	,098	
Group 1	-1,081	,747	2,092	1	,148	,339
Group 2	-1,705	,737	5,349	1	**,021**	,182
Group 3	-,611	,796	,589	1	,443	,543
Constant	1,097	,247	19,769	1	,000	2,994

Interestingly the participants from Group 4 had the highest chances to stay in the additional time. This suggests the interaction effect between the information about the customer they had and information about themselves they had to provide. We believe that the interaction effect was the factor, which made the chances of staying stronger compared to the groups with only one type of information.

5.8.3. Analysis of the output in the additional time

Next we try to understand why some participants stayed in the additional time while others didn't. In order to do that we first check whether there were any differences in the level of output reached in the main time between those participants who stayed and did not stay. Non-parametric test of the equality of means was significant only for Group 1 where we observed a significant difference in preciseness between the participants who stayed and those who did not stay: those who stayed had lower preciseness (p<0,05) and lower speed of typing (p<0,1) in the main time (Table 64). [70]

Thus, in Group 1 those participants who were slower and less precise in their transcripts (which in this group can be generally associated with lower task proficiency) anticipated this situation and tried to compensate their weaknesses by staying in the unpaid time.

Table 64

Comparison of means for output parameters in the main time between those who stayed and did not stay in Group 1
(Welch and Brown-Forsythe robust non-parametric tests of equality of means)[71]

Variable	Statistic[a]	df1	df2	Sig.
Preciseness	5,887	1	24,244	**,023**
Mistakes	,144	1	10,974	,711
Speed_Minutes	2,157	1	11,689	,168
Speed_Words	4,143	1	9,979	**,069**

We can conclude that for other groups neither the initial task proficiency nor the final output quantity and quality were the critical factors in the decision to stay.

Further we analyse whether individual differences in empathy (both trait- and state-like) and beliefs about normative behavior of others could influence the decision to stay. While empathic trait was not a significant predictor of staying, the empathy as state was (Figure 15): those who stayed were the ones who felt much more empathic (independent-samples Mann-Whitney U Test z=2,878, p<0,01), than those who did not stay, but they were not more distressed (z=-0,800, p>0,1).

---

[70] Independent-samples Mann-Whitney U Test also showed that in Group 1 those participants who stayed were slower both in speed of transcribing (z=-2,071, p=0,038) and in speed of typing (z=-1,858, p=,063).
[71] Both test showed the same results for all output variables.

Figure 15

**Comparison of levels of Empathic state and Distress state between those
who stayed and did not stay**
(Independent-samples Mann-Whitney U Test)

Empathic state **Distress state**
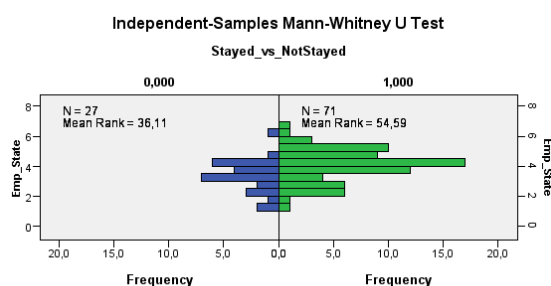
Total N	98
Mann-Whitney U	1.320,000
Wilcoxon W	3.876,000
Test Statistic	1.320,000
Standard Error	125,624
Standardized Test Statistic	2,878
Asymptotic Sig. (2-sided test)	,004
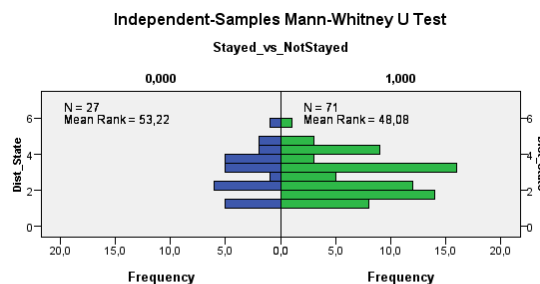
Total N	98
Mann-Whitney U	858,000
Wilcoxon W	3.414,000
Test Statistic	858,000
Standard Error	125,611
Standardized Test Statistic	-,800
Asymptotic Sig. (2-sided test)	,424

If we analyse this relationship within treatment groups we can see that the result was driven by the situation in Group 3, where those who stayed reported much higher empathic state (according to the Independent-samples Mann-Whitney U Test, z=2,501, p<0,01).

Further we have a look at the relationship between the decision to stay and the answer to the question whether people should help other people with disabilities (a proxy for the strength of the prescriptive norm) we can see that in all groups, except Group 2, higher scores for the prescriptive norm were associated with higher chances of staying in the additional time (Table 65). In Group 2, though, this factor did not play a significant role.

Thus, the participants of Group 1 took the decision to stay in the additional time on the basis of their previous performance in the main time; participants of Group 3 who took the decision to stay perceived themselves as feeling more empathic; participants of Group 4 (as also of Groups 1 and 3) had higher chances of staying if they adhered to the prescriptive norm of helping; and in Group 2 the decision to stay did not depend significantly on any of the above factors.

The general conclusion for the above discussion is that when employees are required to open their identity to the customer they are more willing to stay and work in the unpaid time. This behavior generally is not driven by differences in either quantity or quality of previous output, but it was

positively related to the feeling of empathy these participants reported and their agreement with prescriptive norm of helping.

Model 2. Parameter estimates
for logistic regression with binary decision variable "Stayed"
as dependent variable (group split)

Group		B	S.E.	Wald	Df	Sig.	Exp(B)
1	Shoud_Help	,844	,409	4,250	1	**,039**	2,326
	Constant	-3,697	2,260	2,676	1	,102	,025
2	Shoud_Help	,175	,513	,116	1	,733	1,191
	Constant	-,941	3,271	,083	1	,774	,390
3	Shoud_Help	1,523	,700	4,728	1	**,030**	4,586
	Constant	-6,351	3,290	3,727	1	,054	,002
4	Shoud_Help	1,363	,678	4,042	1	**,044**	3,908
	Constant	-5,810	3,747	2,404	1	,121	,003

However, we anticipate that the mere fact of staying did not mean that participants actually did much to improve the existing transcript. That is why further we look at the absolute level of output reached by the participants within the additional time and test, whether there are significant differences in the focus between groups.

If we compare the change in the level of output variables among groups from the end of the main time to the end of the final time (only for those participants who stayed in the additional time) we observe significant changes in such output variables as the amount of recording transcribed ($p<0,1$) and the number of words typed ($p<0,1$).

In particular, the comparisons for separate groups (Table 66) show that participants of all groups, except Group 1, significantly increased the amount of words typed ($p<0,05$). However, in absolute terms this difference is the highest for Group 2 (on average 177 words typed within the additional time). There is also an increase in the amount transcribed, but it is less significant ($p<0,l$) in all groups, except Group 4 ($p<0,05$).

In terms of the variables characterising the quality of output, besides the number of mistakes, we introduce one more parameter – number of paragraphs for 100 typed words. We believe that the introduction of this parameter as an indicator of both general effort and focus in the additional time is justified due to the explicit focus at this part of the experimental procedure on the quality of output, in particular formatting and paragraphing.

While all groups improved accuracy of their transcripts in the additional time (corrected mistakes), only in Group 2 the reduction in the number of mistakes was significant (p<0.05). At the same time explicit focus on formatting led to the significant increase in the number of paragraphs (per 100 typed words) in all groups (p<0.01) except Group 4 (Table 66).

Table 66

Groups split for change in output parameters in the additional time
(main time - final time)
(Related samples ANOVA, bootstrapped)[72]

Group		Mean	Bootstrap[a]				
			Bias	Std. Error	Sig. (2-tailed)	BCa 95% CI	
						Lower	Upper
1	Words typed	-102,105	,555	37,090	0,105	-194,275	-43,240
	§§ in 100 Words	**-,366682**	**-,002373**	**0,083**	**0,001**	**-0,550**	**-0,214**
	Mistakes in 100 Words	,95825	,00137	0,404	0,132	0,318	1,824
	Minutes Recording_Transcribed	-1,20737	,00564	0,416	0,059	-2,198	-0,534
2	**Words typed**	**-177,000**	**-,296**	**41,063**	**0,007**	**-252,449**	**-100,900**
	§§ in 100 Words	**-,598216**	**,007036**	**0,133**	**0,003**	**-0,877**	**-0,312**
	Mistakes in 100 Words	**,60179**	**-,00166**	**0,175**	**0,015**	**0,282**	**0,966**
	Minutes Recording_Transcribed	-3,10786	,01321	1,045	0,064	-5,680	-1,254
3	**Words typed**	**-96,895**	**-1,041**	**31,426**	**0,034**	**-159,818**	**-43,355**
	§§ in 100 Words	**-,419571**	**-,003241**	**0,098**	**0,005**	**-0,625**	**-0,242**
	Mistakes in 100 Words	,58738	,00228	0,321	0,317	0,160	1,250
	Minutes Recording_Transcribed	-1,14579	,00798[b]	,45452[b]	,068[b]	-2,12527[b]	-,32533[b]
4	**Words typed**	**-129,286**	**-2,401**	**31,136**	**0,005**	**-190,446**	**-79,143**
	§§ in 100 Words	-,343174	-,003425	0,125	0,241	-0,635	-0,168
	Mistakes in 100 Words	,45576	,00276	0,230	0,129	0,113	0,909
	Minutes Recording_Transcribed	**-1,76476**	**-,04027**	**0,563**	**0,025**	**-2,950**	**-0,842**

Analysis of the change in the output in the additional time brings us to the following conclusions:

1.   Group 1 was mostly focused on introducing paragraphing (improving formatting of the transcript), which is not surprising if we remember that this was the group where participants most consequently followed the instructions, also in the main time.

2.   Group 2 was the only group which managed to increase significantly all parameters of output.

---

[72] Negative means reflect the increase in the parameter in the additional time compared to the main time; positive means – decrease in the parameter.

3.   Group 3 increased quantity of output and introduced paragraphing, but did not improve significantly in terms of accuracy (we argue that the previous focus on minimizing the mistakes allowed this group to do it within the main time).

4.   Group 4 focused on improving the quantity of output. In absolute terms these improvements were higher than in Group 3, but lower than in Group 2.

Further the non-parametric analysis (Independent-Samples Kruskal-Wallis Test) supports the results of the related samples test: participants from Group 2 transcribed more recording than participants of the other groups, especially Groups 1 and 3 (Figure 16). Participants of Group 4 were also slightly better at this parameter which argues in favour of the proposition that those participants who were confronted with the information about particular customer, if stayed in the additional time focused on further increasing quantity of output. A very similar picture can be observed if we compare the change in the amount of words typed: Group 2 had the highest difference in the amount typed and Group 3 – the lowest.

Figure 16

Box-plots for differences in Amount of Recording Transcribed and Number of Words Typed in the additional time
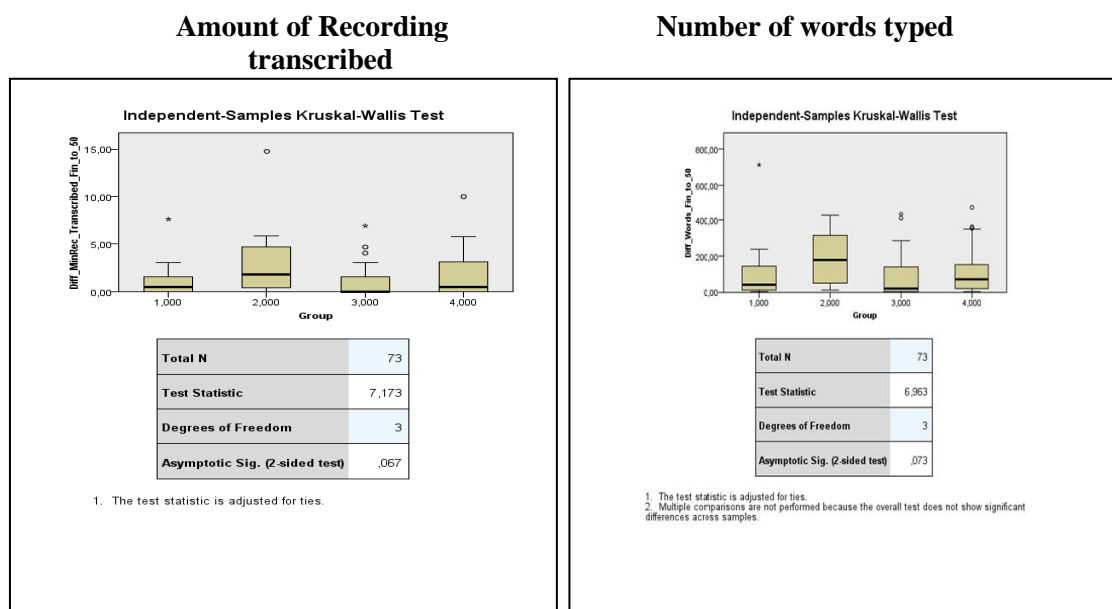
Besides main output variables used throughout the analysis we also introduced several "secondary" variables which allowed us a deeper understanding of the changes made by the participants in the additional time. In particular we compare the transcript made within the main time and further as a result of input in the additional time using the function in MS Word "Compare texts". As a result we come up with a set of four additional change variables: "inserted" (how many pieces of text were inserted), "deleted" (how many pieces of text were deleted), "formatted" (how many formatting

changes have been introduced) and a parameter "changed", which sums up all above changes. It is important to remember that here the unit is a piece of text which has been introduced or deleted at once at some point within additional time.

In this part of the analysis we combine groups according to the principle of anonymity of their employees. Nonparametric comparisons between groups (Independent-samples Mann-Whitney U Test) for groups with and without the requirement for non-anonymity of the employee shows that the first ones (Groups 3 and 4) inserted significantly less pieces of text ($p<0,01$), but formatted significantly more ($p<0,05$) (Figure 17).

Number pieces of text inserted and formatted in the additional time
(Independent samples Mann-Whitney U Test)

	Inserted		**Formatted**	
Total N	73		Total N	73
Mann-Whitney U	365,000		Mann-Whitney U	826,500
Wilcoxon W	1.185,000		Wilcoxon W	1.646,500
Test Statistic	365,000		Test Statistic	826,500
Standard Error	89,776		Standard Error	80,739
Standardized Test Statistic	-3,286		Standardized Test Statistic	2,062
Asymptotic Sig. (2-sided test)	,001		Asymptotic Sig. (2-sided test)	,039

Thus, we can conclude that participants who were asked to open their identity to the customer introduced less content-related changes and more format-related changes.

Participants with the requirement to provide the information about themselves, especially those without information about the customer, stayed more often, but if stayed focused on changes in terms of text formatting, in other words improving the visible aspects of output. This supports our previous conclusion that opening employee's identity motivates self-image-focus: both the fact of staying in the additional time and improving the visible quality of the output can be interpreted in terms of the employee's desire to signal the effort.

Participants with information about the customer, but without the requirement to provide the information about themselves stayed less often, but, if stayed focused on content changes: they typed more words and transcribed more recording. They also worked on improving accuracy, but were far

behind the groups required to open their identity to the customer in terms of formatting. This brings us to the conclusion that information about the customer motivates customer-need-focus, where participants try to improve both quantitative and qualitative parameters of output.

Combination of treatment effects (in Group 4) resulted in the highest motivation to stay and in the quantity-oriented approach towards the distribution of effort. We argue that the participants both exposed to the information about the individual customer and also required to open their identity to this customer were using the opportunity offered in the additional time to provide the customer with those parameters of output that he particularly needed.

5.9. Discussion of the results of the follow-up study

5.9.1. General description of the follow-up study

We conducted a follow-up study in April 2014. The purpose of the study was twofold. On one hand we wanted to uncover the motives behind the behavior of the participants through direct questions regarding importance of different parameters of output for them. On the other hand we wanted to have an "objective" evaluation of the transcripts by individuals placed in the role of customers. In other words we wanted to see whether objective parameters of quality of output correlate with perceived quality of the transcript.

We invited the participants of the original experiment for this study. The decision to use the sub-sample of the initial sample was caused by the consideration of relative complexity of the initial design. We believed that people who did not perform the task themselves could experience difficulty in evaluating the relative importance of the output parameters in a multiattribute task.

We sent an Announcement for participation in the follow-up study via emails (Appendix __) and 35 participants of the original experiment showed interest and took part in the study. Data for 33 participants can be used for comparative analysis of parameter importance. The instructions for the follow up study are provided in the Appendixes 9 and 10.

This study was designed with the purpose of providing us with the answers to the following sets of questions:

1.   Did participants assign different importance to different parameters of output while performing a task in the main study? Are participants aware of the difference in relative importance of output quality and quantity for participants from different treatments?

We know from the results of the main analysis that upon receiving the treatment information participants in different groups paid attention to different parameters of output and also achieved significantly different results. In the follow-up study we attempted, by means of the direct questioning, to elicit the initial opinions regarding the importance of different parameters and compare them among groups.

In order to answer the first set of questions we asked participants to quantify the importance of each output parameter for them (in other words, the participants had to assign weights to four different parameters of output).[73] As a second step we informed the participants about the initial differences in instructions between groups and asked them, with this information in mind, to quantify the relative importance of different output parameters for their colleagues from their and other treatments. In this

---

[73] We proposed to scale the importance (weights) in percent in order to make the procedure more straightforward and easy to grasp.

step participants were explicitly motivated to put themselves into the situation of others and think how they would behave being in their place.

2.   Do the transcripts of different groups differ in the average perceived quality (as evaluated by the general customer)?

While designing the initial study we chose a specific multiattribute task with four main parameters of output. We interpreted the focus on each of these parameters as reflecting either desire to benefit the customer or to satisfy image-related concerns. For example, the extensive focus on preciseness in Group 1 was signalling the conformity with the requirement of the employer, but, as we argued, was not necessary for the customer. The same was true for the accuracy of the transcripts provided by Group 3. More recording transcribed and words typed would be more beneficial for the customer, we argued, than perfectly flawless short text. While the results of the main experimental study were fully in line with our initial argumentation, in the follow-up study we wanted to see, what the final customer would have to say. In particular, it was interesting to know which combinations of quantity and quality, provided by which treatment group (on average) would be evaluated as the best.

For answering this question we first requested each participant to read through several transcripts and evaluate them for their preciseness, accuracy, quality of formatting, etc.[74] Afterwards we asked participants to choose one transcript in each set that they would chose if they could chose as customers. Since each set of randomly combined transcripts was evaluated by three participants, as a result we had an opportunity to say transcripts provided by which group had the highest chances to be chosen as the best.

3.   Does getting into customer's shoes change relative importance of quality and quantity parameters of output?

In the main experiment we argued that exposure to the information about individual customer leads to higher propensity to identify with this customer and the ability to put oneself into his shoes. In the follow-up study we let the participants to evaluate the transcripts as if they were customers. This could encourage them to think about the real needs of the customer and to re-evaluate their own views regarding the importance of different output parameters.

5.9.2. Relative importance of the different parameters of output

The follow-up study was organised in two parts. In the first part, after reminding the participants about the content of the original experiment, we asked them to state how important each of the four main

---

[74] Each participant could evaluate from 1 to 3 sets of transcripts (4 transcripts in each set). The evaluation of each set of transcripts was linked to the payment, so that participants were paid for their effort.

output parameters were for them (Table 67). The focus on *relative* importance of parameters was strengthened by the requirement to assure that the sum for all four parameters should equal 100 percent.

<div align="right">Table 67</div>

Relative importance of different parameters of output

<div align="center">(Table offered to the participants)</div>

Parameter	Relative importance of parameters (in %)
Amount of the recording transcribed ("transcribe as much as possible")	
Closeness of the transcript to the recorded lecture (word-for-word transcript)	
Accuracy of typing (text free from misspellings, usage of capital letters, punctuation)	
Formatting (paragraphing, etc.)	
	100%

The mean values and their range for the relative importance of main parameters of output are given in the Table 68. Participants on average assigned similar importance to the parameters of the amount of recording transcribed and the preciseness of the transcript (around 30 percent out of 100 percent for each). Accuracy of typing at 23 percent was perceived as less important and the quality of formatting, with its importance of about 14 percent, was the last in the list. The spread and standard deviations shown in the Table 68 reflect, however, a significant dispersion in opinions regarding the importance of parameters for different participants.

<div align="right">Table 68</div>

Mean values for perceived importance of different parameters of output

Parameter	Mean relative importance of parameters (in %) (Std.Deviation in brackets)	Minimum	Maximum
Amount of the recording transcribed ("transcribe as much as possible")	30.3 (12.4)	5	50
Closeness of the transcript to the recorded lecture (word-for-word transcript)	32.5 (11.3)	10	60
Accuracy of typing (text free from misspellings, usage of capital letters, punctuation)	23.3 (8.3)	10	40
Formatting (paragraphing, etc.)	13.6 (7.2)	3	30

That is why we further split the dataset with values of relative importance of the output according to the treatment groups to which its participants belonged in the original experiment.[75] Table 69 shows that participants of Group 4 assigned higher value to the importance of preciseness and lower value to the importance of maximising the amount of the recording transcribed, than participants of all other groups. If in Groups 1, 3 and 4 the average values assigned to the parameter of preciseness vary from 34 to 38 percent, then in Group 2 this parameter reaches only slightly more than 25 percent; at the same time the importance assigned to the parameter of output quantity ranged from 30 to 36 percent in Groups 1, 2 and 3 and averaged to 18 percent in Group 4. Importantly, the strongest this difference was between Group 2 and Group 4.

Table 69

Descriptive statistics for responses regarding perceived importance
of different parameters of output (group split)

Group	Parameter importance	N	Minimum	Maximum	Mean	Std.
1	Amount of the recording transcribed	9	10	45	30,00	15,207
	Preciseness of the transcript	9	15	40	34,11	8,852
	Accuracy of the transcript	9	10	40	21,67	10,607
	Formatting of the transcript	9	3	30	13,11	10,822
2	Amount of the recording transcribed	9	20	50	36,11	11,118
	Preciseness of the transcript	9	10	30	25,56	7,265
	Accuracy of the transcript	9	15	35	25,00	6,614
	Formatting of the transcript	9	10	20	13,33	3,536
3	Amount of the recording transcribed	10	15	40	31,50	7,091
	Preciseness of the transcript	10	30	40	34,50	4,378
	Accuracy of the transcript	10	20	30	21,50	3,375
	Formatting of the transcript	10	10	25	12,50	5,401
4	Amount of the recording transcribed	5	5	30	18,00	11,511
	Preciseness of the transcript	5	10	60	38,00	23,611
	Accuracy of the transcript	5	10	40	27,00	13,038
	Formatting of the transcript	5	10	30	17,00	8,367

---

[75] Subjects pool in the follow-up study was limited to the participants of the initial study, who were free to decide whether to take part in this part of the research. Since only 35 percent of the original sample agreed to participate in the follow-up study, we were limited in our ability to control the proportion of subjects from different original treatments. However, we can state that the composition of the sample generally allowed for the balanced distribution of the number of participants among the groups: in particular we had almost equal number of representatives from initial Groups 1, 2 and 3. The only exception is Group 4, which was relatively underrepresented in out follow-up sample, with 5 participants.

In other words employees having the information about particular beneficiary were inclined to value quantity over quality, while employees provided with such information but also required to open their identity to the customer, significantly changed their preferences and moved focus from quantity to quality of output. The difference in the perceived importance of different parameters between Group 2 and Group 4 hints at a strong interaction effect of treatment factors: while the requirement of non-anonymity by itself does not change the relative importance of parameters (Groups 1 and 3 have similar perceptions), information about the customer increases the importance of quantity compared to the general customer condition; and the combination of two factors strongly reduces the importance of quantity in favour of quality.

In order to test whether these differences are significant we conducted a non-parametric Kruskal-Wallis Test, which revealed that there was indeed a significant difference ($p<0,1$) in importance assigned to the quantity of output and the preciseness of output by the participants from different treatments. Figure 18 shows that Group 4 assigned much lower importance to quantity of output and much higher importance to quality of output than Group 2.

Figure 18

Box-plots for relative importance of Amount of Recording Transcribed and Preciseness (groups split)
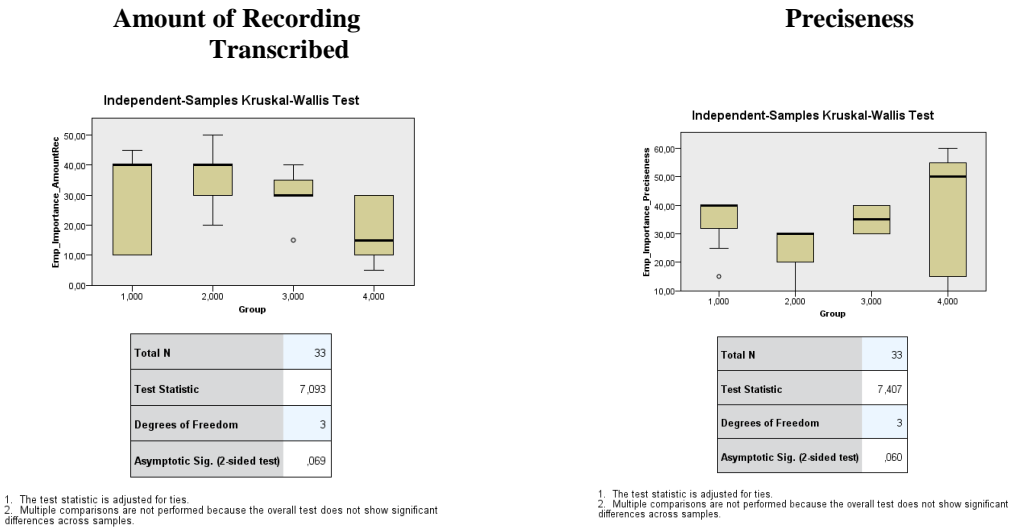
T-test for the significance of the difference in importance between the amount of recording transcribed and the preciseness supports the above conclusions, as compared to other groups, participants of Group 4 assigned less importance to the quantity of output and more importance to the preciseness. The most significant this difference was for Group 2 at $p<0,01$ for the amount of the recording transcribed and at $p<0,05$ for the preciseness of the transcript (Table 70).

Table 70

T-test for differences in relative importance of Amount of Recording Transcribed and Preciseness between Group 4 and other groups

Dependent Variable	I	J	Mean Difference (I-J)	Std. Error	Sig.
Importance of maximizing the Amount of the recording transcribed for the employee	4	1	-12,000	6,403	,071
		2	**-18,111**[*]	6,403	**,008**
		3	**-13,500**[*]	6,287	**,040**
Importance of maximizing the Preciseness for the employee	4	1	3,889	6,085	,528
		2	**12,444**	6,085	**,050**
		3	3,500	5,975	,563

5.9.3. Discovering common beliefs

The next part of the follow-up study was aimed at the elicitation of the participants' beliefs about the behavior of their colleagues both belonging to the same treatment, and to the other treatments. In order to form the beliefs regarding the behavior of others, the participants of the follow-up study had to be exposed to the same information as the one provided to the participants of different treatments in the original experiment. By that time each participant knew only the instructions for her original group and was unaware of the differences in the instructions. We provided the participants with the summary of the initial differences in the information regarding the identification of the customer and the employee herself. After such "debriefing" participants of the study had to fill in the table (Table 71) with values for relative importance of different output parameters for the employees of both their and other treatment groups.

We believed that it was important to assure that participants took this task seriously and were motivated to provide the best possible evaluations. In order to motivate participants we introduced an opportunity of winning a prize for providing the evaluation maximally close to the average statistical results of the original experiment. We suggest that such motivation should have woken up a game spirit on the one hand, while hinting at the very possibility of the differences between groups on the other.

At the next stage in order to make the statistical analysis and the interpretation of the results more straightforward we combined the groups using the requirement of non-anonymity of the employee as the guiding principle.[76] As a result we obtained the average values for the importance of different parameters of output for different treatment groups separately from the employees who were initially anonymous (belonged to Groups 1 and 2) and those who were non-anonymous (belonged to Groups 3 and 4).

---

[76] Previous analysis showed that the requirement of non-anonymity of the employee was a strong predictive factor for the differences in various parameters of output.

Table 71

Relative importance of different parameters of output for different groups of participants
(Instructions for the follow-up study)

Parameter	Relative importance (in %) of output parameters for the group of participants with following instructions*			
	No info about particular customer Name is not provided to the customer	**Have info about particular customer** Name is not provided to the customer y	No info about particular customer **Asked to provide name to the customer**	**Have info about particular customer Asked to provide name to the customer**
Amount of the recording transcribed ("transcribe as much as possible")				
Closeness of the transcript to the recorded lecture (word-for-word transcript)				
Accuracy of typing (text free from misspellings, usage of capital letters, punctuation)				
Formatting (paragraphing, etc.)				
	100%	**100%**	**100%**	**100%**

First we discuss the opinions of the anonymous participants regarding the behavior of both other anonymous employees and non-anonymous employees.

Data in the Table 72 shows that the employees who worked initially anonymously (Groups 1 and 2) believed that, compared to their groups, non-anonymous participants on average assigned less importance to quantity of output (29 percent versus 38 percent) and more importance to its quality in terms of accuracy (24 percent for non-anonymous employees versus 17 for anonymous ones).

Thus, the employees who were not required to open their identity in the original study believed that the requirement of non-anonymity moved focus towards improving the quality of output in its "visible" part, namely the typing accuracy. This result supports our previous conclusions regarding the link between the requirement of employee's non-anonymity and focus on reduction of the number of mistakes in the transcript.

Table 72

**Beliefs about relative importance of output parameters
elicited from anonymous participants (Groups 1 and 2)**

Parameter	Anonymous participants		Non-anonymous participants	
	N	Mean	N	Mean
Amount of recording transcribed	18	38,25	15	28,83
Preciseness	18	30,52	15	33,50
Accuracy	18	17,39	15	23,51
Formatting	18	13,59	15	13,50

Table 73 shows the average values assigned by the initially non-anonymous employees to the importance of different output parameters for other participants either belonging to the groups without the requirement of non-anonymous or with such requirement. We can conclude that the employees from Groups 3 and 4 believed that their groups attached less importance to the amount of the recording transcribed and more to the parameters of quality (preciseness and even more accuracy): They suggested that the relative importance of maximizing of the amount transcribed was 33 percent for the anonymous employees versus 26 for non-anonymous ones, while the importance of minimizing the mistakes for the anonymous employees was about 22 percent versus 27 percent for the non-anonymous ones.

Table 73

**Beliefs about relative importance of output parameters
elicited from non-anonymous participants (Groups 3 and 4)**

Relative importance of the parameter of	For anonymous participants		For non-anonymous participants	
	N	Mean	N	Mean
Amount of recording transcribed	18	33,28	15	26,17
Preciseness	18	28,11	15	32,07
Accuracy	18	22,25	15	27,10
Formatting	18	16,00	15	14,87

Interestingly both the participants from the groups with and without the non-anonymity requirement (rightly) guessed that the requirement of non-anonymity moved attention from quantitative to qualitative parameters of output. This means that the differences observed in the previous statistical analysis based on the experimental data reflect not only the particular pattern of behavior in this experimental setting, but also that these treatment effects (in particular the effect of non-anonymity of the employee) are fully anticipated by the people as general behavioral phenomena.

Additionally we directly asked the participants whether, the information about the individual customer or the requirement to open their identity to the customer would change, first, their general level of effort and, second, the level of effort of the other participants. The question was formulated as close ended both for themselves and for the others and required a choice among three possible answers: "increase in the effort", "do not change he effort:, and "decrease the effort"[77].

The answers for both questions for all participants were positive: in other words they believed that both the exposure to the information about the individual customer and the requirement to open their own identity would result in the increase in effort. The only exception from this pattern was observed in the opinions of the participants from Group 3, who believed that their effort was not influenced by the necessity to open their identity. While the employees from other groups anticipated the potential effect of non-anonymity on their effort level and effort of others, participants from the groups initially exposed to this effect in the experiment, explicitly did not (want to) attribute their effort to the influence of the treatment factor. This situation is illustrated by the Table 74, where we can see that the t-test confirms the significant difference in the opinion about potential change in behaviour due to the introduction of non-anonymity between Group 3 and other groups. The most significant this difference is between Groups 3 and 2 (p<0,01).

Table 74

T-test for differences in beliefs about change in effort due to the requirement of non-anonymity between Group 3 and other groups

Dependent Variable	Group (I)	Groups (J)	Mean Difference (I-J)	Std. Error	Sig.
Change in Employee's effort if required to provide her name to the customer	3	1	-,367	,205	,085
		2	**-,589**[*]	,205	,008
		4	**-,500**	,245	,050

*The mean difference is significant at the 0.05 level.

5.9.4. In customer's shoes

Finally, after participants had a chance to read the transcripts of others and to evaluate them (results of this part are discussed later), they were asked to think about the importance of different parameters of output for the customer. Conducting the related samples Wilcoxon Signed rank test for changes in importance of output parameters as evaluated by the participants before and after being put into the "customer's shoes" (Table 75) we come to the conclusion that the participants significantly increased

---

[77] For the statistical purposes the positive answers were coded as "1", neutral answers as "0" and negative answers as "-1".

the relative importance of qualitative parameters of output -preciseness and formatting (with p<0,01 and p<0,05).

Test for difference in relative importance of different output parameters from customer's and employee's perspective
(Wilcoxon Signed Ranks Test)

	Amount of Recording (Importance for Customer – Importance for Employee)	Preciseness (Importance for Customer – Importance for Employee)	Accuracy (Importance for Customer – Importance for Employee)	Formatting (Importance for Customer – Importance for Employee)
Z	-,282[b]	-2,326[b]	-,202[c]	-3,365[c]
As. Sig. (2-tailed)	,778	,020	,840	,001

b. Based on positive ranks.

c. Based on negative ranks.

If we look at the changes in the evaluations within the individual groups we can state that the importance of preciseness diminished in all groups, but the strongest result was in Group 1 (Figure 19). This is not surprising if we remember that this group was the one most closely following the requirement for preciseness set in the instructions.

Relative importance of preciseness from employee's / customer's perspective
(groups split)
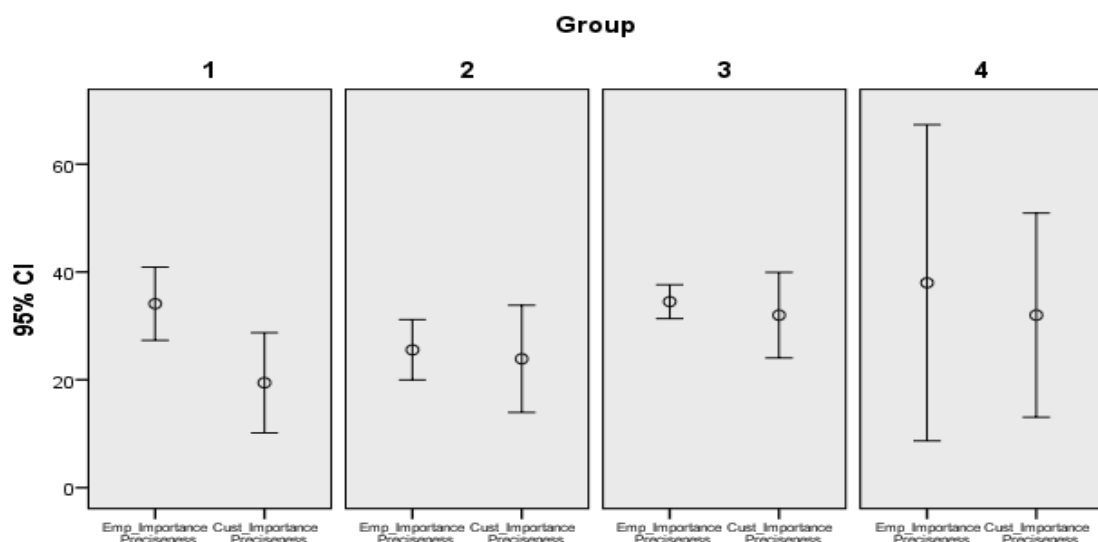
However, after reading and comparing the transcripts, the employees could form their own opinion regarding the comparative importance of complete preciseness. The significant change in the relative importance of preciseness signals, in our opinion, the realisation that from the customer's perspective

the perfect preciseness, in the presence of other factors, is not necessarily the prerequisite for the good transcript.

Another results of this analysis is that all groups at this stage became united in the opinion that formatting of the text has higher importance than they previously believed. However, as we can see from the Figure 20, Group 1 showed the strongest change in their opinion regarding the relative importance of this parameter.[78]

Figure 20

Relative importance of Formatting from employee's / customer's perspective (group split)
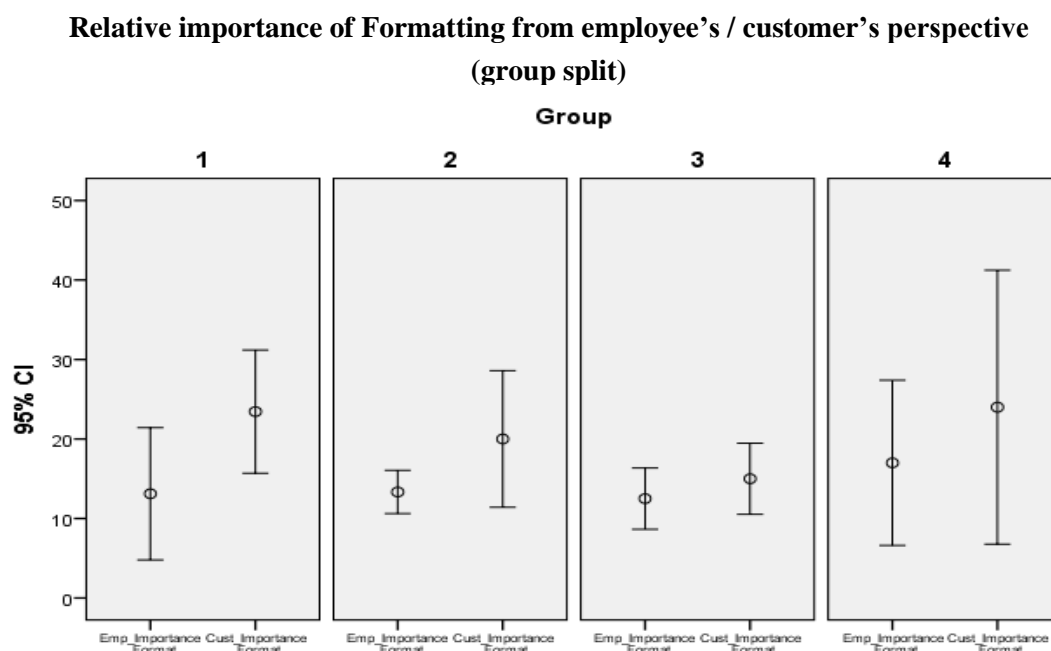
If we now (after "putting the employees into customer's shoes") analyse the differences among the groups in their opinions regarding the relative importance of different parameters of output for themselves, we can see that the picture still resembles the one which we observed in the beginning (Table 76). In particular, participants of Group 4 still assigned less importance to maximising the amount of the recording transcribed (though, at lower significance). At the same time the importance of preciseness was re-evaluated by the groups, especially by Group 1, where the opportunity to see the output from the customer perspective led to the decrease in importance of preciseness, which as a result became significantly less important than in Groups 3 and 4 (p<0,05).

---

[78] None of the groups changed significantly their opinion about the importance of maximizing the amount of the recording transcribed.

Table 76

T-test for differences in relative importance of Amount of Recording Transcribed and Preciseness from customer's perspective among groups

Dependent Variable	Group (I)	Group (J)	Mean Difference (I-J)	Std. Error	Sig.
Perceived importance of the amount recording transcribed for the customer	4	1	-12,000	7,394	,115
		2	-14,778	7,394	,055
		3	-14,000	7,260	,064
Perceived importance of the preciseness for the customer	1	2	-4,444	5,905	,458
		3	**-12,556**[*]	5,756	,037
		4	-12,556	6,987	,083

*The mean difference is significant at the 0.05 level.

There are two major conclusions form this part of the analysis:

- Participants from all groups anticipate that non-anonymous employees assigned less importance to the amount of the recording transcribed and its preciseness and more importance to the accuracy of the transcript and formatting, than anonymous participants;

- After being involved into the evaluation of the transcripts, all participants downscale the importance of preciseness and increased the importance of formatting, but this change is the most significant in Group 1, which participants significantly increased the importance of both formatting and accuracy.

5.9.5. Evaluation of the transcripts

Further we move to the discussion of the results for the transcripts evaluations made by the participants of the follow-up study. As it was described above, each set of transcripts was independently evaluated by three participants. Each evaluator had to grade the transcript based on five major characteristics: easiness to read, preciseness, accuracy, formatting and general quality. Each parameter was measured using the 7-point Likert scale ranging from "Very good" to "Not at all".

First of all we check the level of agreement between evaluators regarding their scoring for each parameter (Table 77). As we can see for 4 out of 5 parameters the level of agreement is quite high (Cronbach's Alpha > 0,7), the only exception is the parameter of preciseness (Cronbach's Alpha= 0,5). Both specific parameters of quality such as accuracy and formatting, and general parameters such as easiness to read and general quality had comparable standards for all evaluators and could be easily assessed. Preciseness, however, could be hardly evaluated without prior listening to the original recording. Hence, evaluation the transcript as precise or not precise was rather subjective, which led to the lower agreement between evaluators. We see this as a support for our previous proposition that preciseness, though being as accuracy the parameter characterising the quality of output, was differing

from accuracy in terms of its "visibility" (ability to be easily evaluated without prior listening to the recording).

Level of agreement (Cronbach's Alpha) among the scores for qualitative parameters of output from three independent evaluators

Parameter	Cronbach's Alpha (N of items =3)
Easy to read	0,734
Preciseness	0,459
Accuracy	0,738
Format	0,730
General quality	0,734

Preciseness

Further we move to the analysis of the comparisons between average scores for the specific and general parameters of quality assigned to the transcripts from different treatments. We start with the discussion of scores for the preciseness.

Comparing the average scores for the preciseness of the transcripts among treatment groups we can see that the transcripts produced by Group 2 enjoyed the highest average scores and also the interquartile range for the scores in this group is the narrowest (Figure 21). This means that Group 2 not only provided the transcripts perceived as highly precise, but also that at least half of its participants (interquartile range) scored from "5" to "6" (on a 7-point Likert scale) for preciseness. Other groups are characterised by the lower average preciseness for their transcripts.

Interestingly the transcripts of Group 1, which according to our previous analysis was the one taking preciseness maximization most seriously and having the highest objective preciseness at the end, was not seen as the most precise by the evaluators.

Both in Group 3 and especially Group 4 the interquartile ranges and the general spreads of scores are larger than in other groups: about a quarter of transcripts in these groups scored lower than "4", while in Group 2 only a negligible proportion of transcripts falls into this category.

Figure 21

Box-plots for average scores for Preciseness of evaluated transcripts

(groups split)
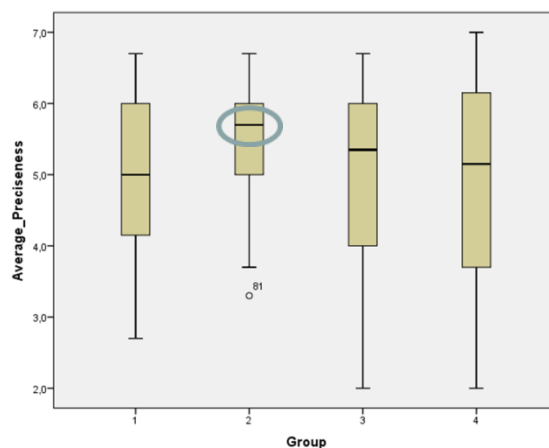
Accuracy of Transcript

Analysis of the scores for perceived accuracy of the transcripts shows the best result for Group 2 which enjoyed the highest average scores and the narrowest range (Figure 22). As in the case of the preciseness, Group 3 and especially Group 4 are characterised by the large interquartile ranges and wide spreads. This means that the transcripts of these groups were rather heterogeneous in terms of quality.

Figure 22

Box-plots for average scores for Accuracy of evaluated transcripts

(groups split)
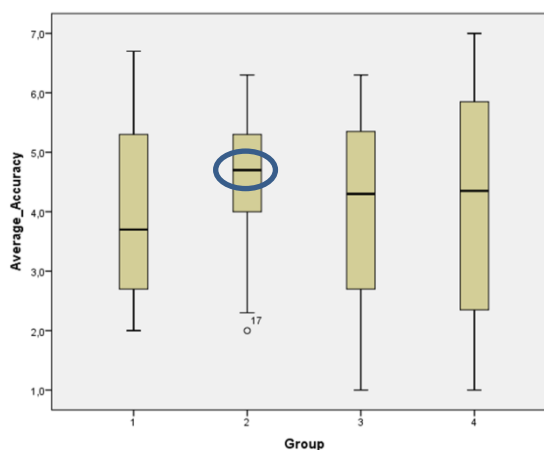
Text formatting

Group 3, however, was the best in terms of formatting (Figure 23), which is another parameter characterizing visible quality, not explicitly mentioned in the instructions within the main part of the experiment. Group 3 also has the smallest range for this parameter (although it is not as small as the interquartile range for parameters of preciseness and accuracy in Group 2).

Box-plots for average scores for the quality of Formatting of evaluated transcripts (group split)
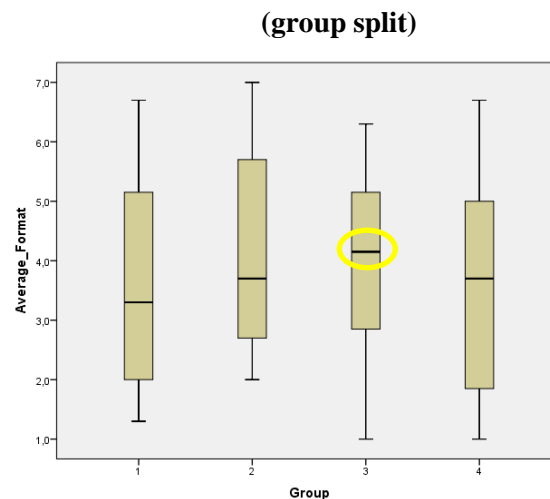
Group 4 has the largest spread for the scores for the quality of formatting among the groups (the transcripts of this group were evaluated as ranging from "1" to "7" in terms of formatting), which means this group is characterized by the wider differences among participants, some of them achieving very high scores for the parameters of quality and some low scores.

General parameters of quality

The above analysis of the separate parameters of quality makes the results for the general parameters not that surprising (Figure 24). In particular, Group 2 has the highest average value for both "easiness to read" and "general quality"; Group 4 has the largest spread of values, especially for interquartile range. If in Group 2 interquartile range for the average scores (among 3 evaluators) for general quality varied from roughly "4" to "5", for Group 4 this range lied between "2,5" and "5".

From the deeper look at the box-plots we can also conclude that the transcripts provided by the participants of the groups exposed to the information about individual customer (Groups 2 and 4) had slightly higher general quality compared to the groups having only general information about the customer (Groups 1 and 3).

Figure 24

Box-plots for average scores for "Easiness to read" and "General quality" of evaluated transcripts (group split)

Easy to read   **General quality**
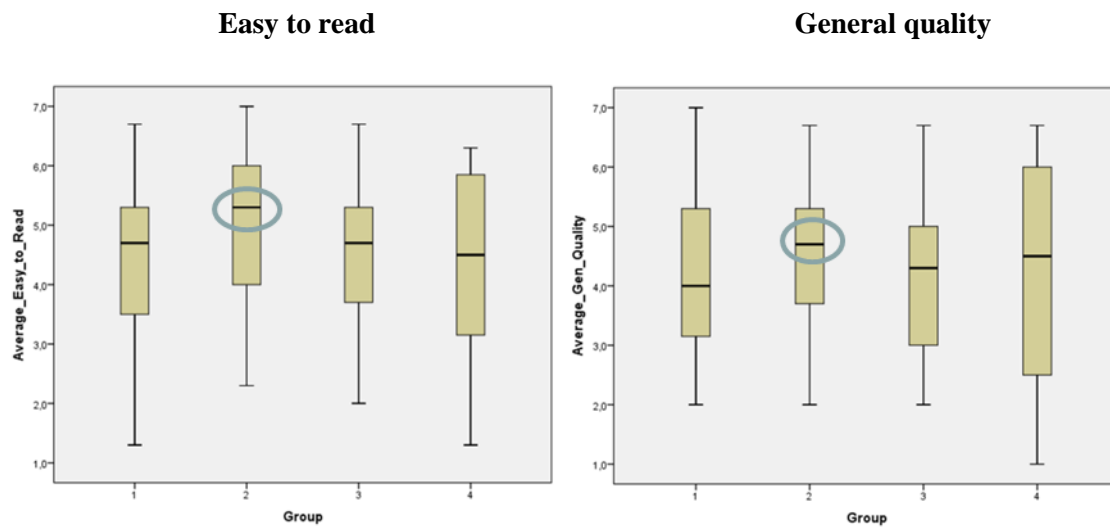
We believe that the analysis of the parameter of general quality provides us with a good picture of customer's preferences in terms of quality. That is why we go one step further in this analysis and instead of average scores for general quality we introduce three dummy variables for low, medium and high quality transcripts and compared the number of across the treatment groups.[79].

Figure 25

Number of high- / medium- / low-quality transcripts (groups split)

---
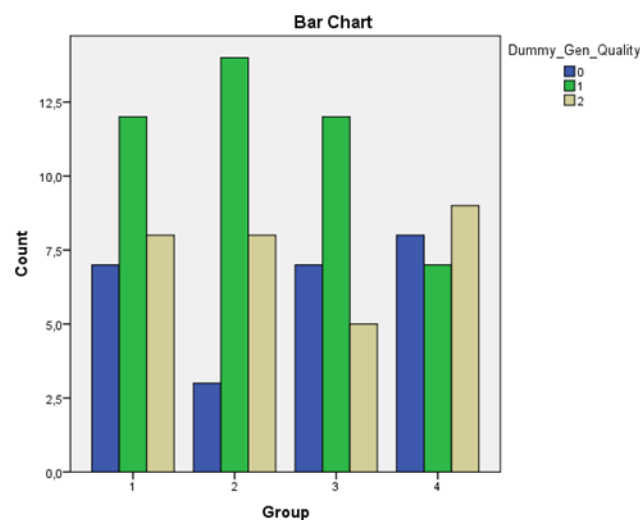
[79] Dummy_Gen_Quality=0 (low quality) if average evaluation lies between 1 and 2,9 points,
Dummy_Gen_Quality=1 (medium quality) if average evaluation lies between 3 and 5 points,
Dummy_Gen_Quality=2 (high quality) if average evaluation lies between 5,1 and 7 points

Figure 25 (above) summarises our results and shows that Group 2 had the lowest number of low quality transcripts (3 transcripts) and the highest number of medium quality transcripts (14 transcripts), while Group 4 had both the highest number of low and high quality transcripts (8 and 9 transcripts respectively). This supports once again our intuition that participants of Group 4 ranged significantly in terms of quality of their transcripts.[80]

Quantity of the transcript

However, not only the quality of transcripts was important for the customer, but also (and probably even foremost) its quantity. In the next step we checked whether the parameters of perceived quality correlated with the final parameter of quantity – number of words typed. We do it both for the whole sample and for the separate groups (Table 78).

What we can see form this analysis is that for aggregated sample all parameters of quality correlate with quantity, except for the parameter of formatting. If we remember that amount of words typed was related to the participants proficiency in the task we can suggest that those participants who were able (and willing) to maximise the quantity had higher chances to provide better quality.

However, if we look at the correlations within groups, we can see that these relationships are significant only in Groups 4 and (to a much lesser extent) in Group 3. Thus, the results for the aggregated sample can be seen as being driven by the correlation in Group 4. This once again supports our conclusion that in Group 4 those participants who provided high quality of output also had high chances of performing well in terms of quantity. Since this group was under the combined influence of both treatment effects, we argue that the all its employees attempted to improve all parameters of output, but due to the initial differences in the task proficiency, some of them achieved significantly better results than others.

---

[80] The difference between Groups 2 and 4 is significant at p<0.10 (Chi-square statistic is 4.646; exact p-value is 0.097).

Table 78

Correlations between final Number of Words Typed in the transcript and scores for quality received by the transcript

		Gen_Q	Easy_to_Read	Preciseness	Accuracy	Format
Words_Fin	r	,415**	,316**	,317**	,385**	,054
	N	100	100	100	100	100

Group			Gen_Q	Easy_to_Read	Preciseness	Accuracy	Format
1	Words_Fin	r	,263	,202	,033	,184	,019
		N	27	27	27	27	27
2	Words_Fin	r	,198	-,040	,180	,192	-,349
		N	25	25	25	25	25
3	Words_Fin	r	,499*	,405*	,286	,351	,186
		N	24	24	24	24	24
4	Words_Fin	r	,568**	,548**	,717**	,673**	,205
		N	24	24	24	24	24
**. Correlation is significant at the 0.01 level (2-tailed).							
*. Correlation is significant at the 0.05 level (2-tailed).							

The quest for the best transcript

Finally we asked each participant which of the four transcripts they would chose if they were the customer. As set of transcripts always consisted of transcripts made by participants from different treatments, we could match the choice of the evaluator to the treatment. Below is the comparison between groups in three graphs (Figure 26). First graph shows how many transcripts in each group were chosen by at least one evaluator as the best transcript and how many were not chosen by any of the evaluators.

For example, 14 transcripts from Group 1 were never chosen as the best and 10 were chosen as the best by at least 1 evaluator. In Group 4 only 9 transcripts were never chosen, while 15 transcripts were chosen at least once.

Next two graphs having the same columns for not-chosen transcripts, splits chosen at least by one evaluator, two evaluators and eventually all three evaluators. It is interesting to see that there is very high agreement between evaluators regarding transcripts from Group 4: nine of the transcripts from this group would be chosen by 2 or all three evaluators. This means these transcripts were good both in terms of quantity and quality and could meet the requirement of different customers. Group 2 has the second best result with only 12 transcripts never chosen and 13 transcripts chosen at least once.

Figure 26

Number of transcripts chosen as the best transcript by

Nobody/ at least
1 evaluator

Nobody/
1 evaluator / at least
2 evaluators

Nobody/
1 evaluator / 2 evaluators /
3 evaluators

5.9.6. Conclusions for the follow-up study

There are several conclusions which can be made based on the above analysis. The most general from them is that by means of the experiment and follow-up study we could confirm that the change in the context in which a particular task is performed could lead to the change in motives which govern employee's behavior, which in its turn resulted in setting the focus on different attributes of output.

In particular in our experimental settings we first introduced a "no-purpose" context for all participants and afterwards randomly assigned them to one of the four different "purpose-laden" contexts: the first one giving only general impression about the purpose of the task and assuring anonymity of the employee, second adding the information about a selected customer; third lifting the requirement of employee anonymity; and the last one combining all previous types of information given and required.

Introducing "purpose" resulted in the general increase in effort, which however was channelled towards different parameters of output depending on the specifics of the context.

In the context of general information (Group 1) we could observe the actions aligned with the requirements set by the employer in the instructions: in particular providing maximally precise text was given a priority over other parameters of quality or quantity of output. This appropriateness of this attitude was, however, re-evaluated after the employees were provided with the opportunity to "get into the customer shoes" through reading and evaluating the transcripts.

When we added information about the selected customer (Group 2) we could observe a move in focus from following the employer's requirements towards meeting "customer-need" (balanced increase in all output parameters with stronger focus on increasing the amount transcribed, as being the output

variable which for objective reasons could not be improved by the customer herself). Results of the follow-up study supported the conclusions of the main study as the transcripts provided by the participants of this group were not only the best in terms of quantity but were also evaluated as the best in terms of quality.

Lifting employee's anonymity without introducing her to the selected customer (Group 3) led to the increased focus on "saving image", in other words on providing the output which was clearly showing the effort of the employee. Since amount transcribed as well as preciseness of the transcript were not parameters which could have been easily evaluated by the customer (especially in case of the person with hearing disabilities unable to hear the recording herself), this group concentrated on providing the least amount of mistakes and best formatting. Such pattern of behavior led to the situation where this group objectively provided the worse result in terms of quantity, but best result in terms of "visible" quality.

Finally combining treatment effects (Group 4) we created a new context where, as it turned out, information about selected individual customer was a factor mitigating the effects of the non-anonymity of the employee. In particular participants of this group provided the transcripts of better general quality and quantity than participants of Group 3, but shorter than Group 2 and less precise than Group 1.

If we consider pure influence of treatment effects (providing information about the selected customer or requiring the employee to open her identity to the customer), providing the employees with information about the individual customer (beyond the general purpose) leads to a balanced effort in terms of quantity and quality, which is not necessarily exactly reflects the requirements set by employer, but at the end orients itself towards the needs of the customer.

Non-anonymity of the employee can have negative effect on the output as it moves the focus towards the parameters of output "signalling the effort", which does not necessarily reflect the real needs of the customer. However, in the combination with the information about the selected customer the negative influence of this effect is mitigated and from the customer's perspective the output provided is characterised by both higher quantity and quality. Interestingly, interaction between treatment effects led, on the one hand, to the better understanding of the complex structure of output and more balanced approach to maximisation of individual parameters, and on the other, to the higher importance of initial task proficiency in determining the general level of output.

VI.   Conclusions and Implications

6.1. General conclusions for each treatment group

To help – Make it easier or possible for (someone) to do something by offering them
one's services or resources; to improve (a situation or problem)
Oxford online Dictionary[81]

6.1.1. Job Meaning as a new motivational mechanism

In scientific literature a lot of attention is paid to the issue of importance of human resources for the success of the organisation. However, according to many researchers the mantra "people first" is often nothing more than a rhetoric statement. According to Calabrese (2012), human resources in present organizations are not fully involved and do not provide all the benefits for the organization, in other words their abilities and potential are partially underutilized. Their contribution in reaching company's missions is limited to performing the required duties and functions, or congruent with incentives they are exposed to. Researchers (Calabrese, 2012) base their conclusions both on the review of managerial literature and on the results of interviews which they conducted in the framework of their study. More than 70 percent of their respondents (197 MBAs students, working in the service sector) showed no full motivation and commitment toward their work. Among the causes the authors name the loss of meaning that both employees and managers feel about their profession.

The dimension of "job meaning" highlighted by Calabrese's respondents is in fact missed in many motivational theories. One of the approaches towards analysing the role of meaning was offered by Frankl (1962), who stated that a lot of people lack meaning in life. People can be motivated by satisfaction of their needs, but at the same time they can be demotivated by psychological boredom, which has its roots in the lack of meaning. Moreover achieving long-lasting satisfaction requires moving beyond individual wellbeing and involves relationship with others. Professional life is not an exception in this case: profession can gain meaning due to realization of the fact that it serves other people. In this case professional motivation is found not in the profession itself, but in the deep purpose of the job.

Calabrese (2012) suggests to analyse motivation in terms of three non-conflicting and self-reinforcing dimensions: functional dimension which considers primary needs (salaries, etc.), relational dimension that considers employee's relationships with colleagues and management, self-esteem etc.) and a meaning dimension. Importantly the researchers believe that each of these dimensions can be activated only if management is focused on their activation. They admit that both in theory and in companies' policies functional and relational dimensions are widely analysed and employed, whereas meaning dimension is normally neglected. They state that if managers help their employees to discover the

---

[81] http://www.oxforddictionaries.com/definition/english/help

meaning of their work, such employees could over-perform in comparison to the employees without such meaning approach to their work.

According to Grant (2009, 2011) the notion of job meaningfulness is closely related to the opportunity to perceive it as aimed at helping others. Based on the meta research combining studies for more than 11 thousand employees from different industries, Grant (2009) concludes that people believe that their work is meaningful only if it has positive impact on others. Respondents in the US, as well as in China and Eastern Europe believe that activity qualifies as meaningful work if it contributes to the society.

If perceived meaningfulness of the job can be seen as the factor increasing employees' motivation, then it is natural to ask which tools can be employed to make the job more meaningful. While in the managerial literature the construct of perceived impact on others emerged as the discussion topic, we pose the question regarding the forces behind this construct. In other words, why do people want their job to have a positive impact on others? Getting an answer to this question, we believe, can be of great importance for designing managerial interventions aimed at increasing motivation through meaning.

In our search for relevant ideas we employed interdisciplinary approach and turned to the literature in economics (especially behavioral and experimental economics), organisational theory, psychology, and sociology. As a result we came up with the set of two generic sources of motivation responsible for enjoying the positive impact on others: empathy and image concerns. Taking this as a starting point we argued that the informational context in which employee is operating could alter the strength of these motivators, and thus, influence performance.

Research in economics shows that information about social context can shape decisions in (at least) three ways: information can be treated differently when there is a social context altogether; individuals can derive utility from other's outcomes and finally, from how they treat others in particular social relationships (Huettel and Kranton, 2012). Research in neuroeconomics says that there are even separate pathways in the brains for processing social and non-social information[82] (Behrents et al, 2008 cited in Huettel and Kranton, 2012).

However, the parameter of social context is also rather complex and "others" in social interaction can be seen as abstract group or perceived as a specific needy individual, while in some job environments they are not considered at all. We saw our task in distinguishing between such environments in order to see whether the complex construct of job impact can be operationalised in economic terms and managerial implications can be drawn.

In order to do that we compared the contexts both deprived of meaning, and also instilled with it. We provided the employees with an opportunity to learn what impact their work will make on the customer. Moreover we distinguished between general customer and individual customer.

---

[82] For example rewards coming from person or computer.

While the behavior in the control time can be seen as the one reflecting "no impact" motivation, the behavior in the main time for all four experimental conditions represents the results of the meaningful work.

We expected that introduction of meaning in form of the general information about the customer would motivate the employees to exert higher effort on behalf of the customer. However, in real life environments where employees have to perform multiple tasks it is important not only to motivate employees' effort, but also to achieve optimal trade-off between different tasks. The trade-off points are often individually decided and under conditions of environmental uncertainty can only hardly be controlled by the management.

This is especially important in not fully specified contracts / tasks, where customer's satisfaction is a primary goal of organization. If tasks are multidimensional then taking decisions about the importance of different parameters of output requires additional effort of the employee.

From agency theory we know that incentivised agents will perform better on easily quantifiable parameters of output (Holmstrom and Milgrom, 1987). Bonuses for less easily quantifiable are recommended. If we add the environmental uncertainty and incomplete contracting condition, it becomes clear that the both the overall level of effort and the decision about the attributes trade-offs are to be left at the discretion of employee. Since monitoring in such cases is quite costly (both in economic and in psychological terms) only customer can perform the ex-post verification of the results, and bear the consequences of employee's choices. That is why it becomes increasingly important to introduce ex-ante mechanisms to motivate the employee to both maximize the amount of effort exerted on the task and also distribute this effort between task attributes in the way benefiting the customer.

We introduced employees to the task which is done for the existing real customers - the students with disabilities – represented by university-based non-profit organization Mitschriften Börse. The choice of customer was intentional and aimed at moving the focus of employees away from the relationship between employer (experimenter in our case) and employee towards the customer. At the same time the specifics of the customer group allowed to have a direct link between employee's work and impact of her effort on the customer.

We start by introducing the employees working for "purposeless" task to the general meaning of their work (in the form of general customer) and ask whether it is enough to motivate employee's effort and moreover to direct it towards willingness to discover and meet the customer needs. Further we provide a short discussion and a summary of the results separately for each treatment group.

6.1.2. General context of socially valuable task: Group 1

In Group 1 we provided employees with the general purpose of their work, thus allowing for higher meaningfulness condition compared to the no-purpose frame. If we compare these two frames we

come to the conclusion that such intervention led to the increase in the overall amount of effort, with the decision regarding the trade-offs between attributes influenced by the explicit (though not monitored) requirements of the management.

As a result of the analysis of the behavior in Group 1 we can argue that by providing the employees with the general information about the customer we can motivate the effort of employees, but not necessarily the focus on meeting customer needs. The participants in this group were focusing their effort on following the instructions provided by the "employer" (in our case the experimenters) without questioning whether these requirements match the real needs of the customer. This pattern is clearly seen by analysing the pattern of change in the parameter of preciseness, characterising the closeness between the recorded speech and transcribed text. While we know that employees in all groups slightly increased preciseness, the employees of Group 1 were the only ones who increased it significantly upon receiving the instructions. As a result the transcripts provided by them are characterised by highest average preciseness among all groups. At the same time the transcripts made by this group contain the highest average amount of misspellings, with the lowest average amount of the recording transcribed among the groups.

In other words employees with the general information about the customer had a clear focus on achieving high preciseness, which was set as the requirement (though not monitored) by the employer. By focusing on preciseness participants avoided the costly decision-making effort regarding the trade-off between the preciseness and any of the other output variables. Moreover there is an overall agreement among the employees regarding the focus of effort: variation in preciseness is related mostly to the differences in language skills - with native German speakers being able to provide more precise transcripts. The same is true for the amount transcribed and speed of typing: the variation in these parameters is also related to task-related skills (which are also reflected in the answers regarding the perceived easiness of the task). Thus, having a common focus on improving preciseness (compared to the "no purpose" control time) participants in this group differed in their output mostly due to their task proficiency (language and typing skills).

Also in the additional time phase those participants from Group 1 who took the decision to stay did not improve significantly in either accuracy or amount of the recording transcribed.

Conclusion

From this analysis we can conclude that the introduction of "pro-social" general task purpose leads to the slight increase in effort compared to the "no purpose" condition. Importantly this effort is concentrated on following the pre-set performance requirements reflecting the conformity with existing rules. However, in case when requirements set by the employer are not optimal in terms of matching customer needs, such pattern of behavior can lead as in our case to the output providing comparatively lower level of customer satisfaction.

In particular, the evaluations of the transcripts in the framework of our follow-up study revealed that the average transcript of this group did not score as the best among four groups in any of the four output parameters. Also as a result of randomised individual selection of the best transcript among four transcripts from each treatment group 17 out of 27 transcripts (the highest proportion of all groups) were never chosen as the best ones.

Managerial perspective

The introduction of the general job purpose (in our case in prosocial context) is beneficial for motivating the employees to exert effort based on the requirement of the management, but it is not sufficient to motivate customer-focused decision-making. Thus, in the tasks with little uncertainty about customer needs and quality-quantity trade-offs introducing general customer can be a way to motivating employees' extra effort. However, if there is uncertainty regarding the optimal trade-off between quality and quantity of output (related for example to the specific customer needs, or task-specific skills of the employees, time constraints, etc.) other mechanisms are necessary to trigger customer-focused thinking and perspective-taking.

6.1.3. Customer gives meaning: Group 2

We assume that any company sees satisfaction of customers' needs as one of its most important goals. In many fields – especially in service sector – employees can and should play the essential role in achieving this goal. In terms of employees motivation this means that management faces two related but still distinct tasks: one is to motivate employees to work (perform her duties in accordance with job description) and another is to motivate employees to "go extra mile" and create a product or provide a service able to satisfy the customer. From previous discussion we can conclude that providing employees with general purpose of their jobs can help in succeeding in the first task, while for the second –we argue – empathic feelings play a significant role.

According to the research by Grant (2014) there are some jobs which people see as high in meaning (for example, fire chief, neurosurgeon) and others which are low in meaning (revenue analyst, web operations coordinator). However, Grant (2008, 2009) believes that almost every job can be made more meaningful. The problem is that while most of the jobs do have impact on others, employees are too distant from the end users of the products and services (for example, automotive safety engineers, and medical scientists). That is why leaders at companies manufacturing farming equipment (e.g. John Deere, US) invite farmers who buy the tractors for the talk with employees, users of Facebook who found there friend meet with software developers, Wells Fargo films videos about customers who got low-interest loans and thus were rescued from debt.

As we hypothesized in the beginning of this Thesis, introduction of the specific individual customer created the feeling of empathic concern and motivated the employee to exert higher effort on behalf of

the customer. Moreover, this effort is characterized by customer-focus, where trade-off between different output parameters depends not only on the employer's self-interest, but also on the perceived customer needs.

The analysis of the literature in the field of social psychology and development leaves no doubts about the fact that empathy, being one of the most important motivators, also boosts altruistic behaviour (Edele et al, 2013).

In our research in Group 2 we added to the general information about the socially needy customer the more personal touch, by exposing the participants to the short story and facial image of a particular customer, for whom they performed the task.

According to our predictions, adding the information about specific customer (in the context of pro-social job purpose) should allow employees to identify with the customer (through the mechanism of empathy), take customer's perspective and focus on output parameters essentially important for the customer. We argue that it is easier to identify with the single customer, than with a group of customers. Such identification should allow on one hand "perspective taking" and on the other "empathic sympathy". Due to the perspective taking the employee can better understand what the customer needs and sympathy should move her in the direction of satisfying these needs.

In our particular task we see the maximisation of the quantity of output as such important parameter with quality standards individually determined by the employee. The reason for that is that taking the decision about quantity / quality trade-off under the time constraint the employee should remember that the beneficiary is unable to hear the recording (thus, exact preciseness is not necessary and some amount of misspellings could be tolerated without complete loss of meaning).

Indeed the employees in Group 2 were the only ones among four groups who had higher chances of increasing the speed of transcribing (measured in minutes of the recording transcribed in one minute of working time) than decreasing it and they were the only ones who significantly increased the speed of typing (measured in the number of words typed in one minute of working time).

It is clear from the statistical results that this group had to make a difficult trade-off between preciseness and speed of transcribing. It is especially important to note that this trade-off was decided on the basis of believes about the helping behaviour of others. In particular, if the employee believed that helping disabled people is a behavioral norm for all people (high agreement with the statement that generally people do help others with disabilities), she provided lower speed of transcribing and higher preciseness; if, on the contrary, the believes about helping behaviour of others were rather low, the employee was focusing on maximising the amount transcribed, sacrificing the preciseness. Thus, we interpret this result in the following way: those participants who saw others as less helping were feeling more responsibility for providing the customer with the output parameters most essential for him. That is why they were trying to transcribe as much as possible. In other words, compared with the participants who scored high on the question regarding the descriptive norm of helping (having

strong believes about helping behavior of others), they were trying to provide the customer with a significantly less precise but longer transcript covering more material on the tape.

Interestingly, while Group 1 was the best in terms of preciseness among all groups, regression analysis showed that if we account for the difference in the beliefs about helping, Group 2 provided the transcripts of with even better preciseness than Group 1.

Employees in this Group 2 were less focused on accuracy – they were the only group who did not decrease the amount of mistakes after getting the instructions. The amount of mistakes in this group was strongly related to the task proficiency of the participants, as well as to the individual level of empathic trait. Importantly, preciseness was also positively related to the empathic trait, while the quantity of the output did not relate to the differences in trait-like empathy. We interpret this result in terms of higher effort provided by more empathic employees in this treatment: they assure better quality of output without letting the quantity suffer.

As the responses to the empathic trait questionnaire give us the measure of perspective taking, we can conclude that these employees were more prone to accepting the perspective of the customer (putting themselves into the customer's shoes) and as a result they worked harder to keep both quantity and quality high.

Thus, adding the information about the customer moved the focus of participants into the direction of maximising quantity, while still making the effort to keep a high level of preciseness. The behavior of the participants of this group is less uniform as it differs depending on their beliefs about the helping behaviour of others. At the same time we argue that the agreement with the statement regarding others' helping behavior (descriptive norm) should be analysed in connection with the experimental context and related to the responsibility for completion of the task. In particular, if participant believed that others do not help people with restricted abilities, than this meant that this is also true for current task. Hence, the person felt responsible for providing as much quantity as possible. If, on the other hand, the others are seen as willing to help, then the employee saw the responsibility for completion of the task as shared and allowed her to invest more in achieving higher quality.

It is interesting that in Group 2 only a little more than half of the participants stayed in the additional time (14 out of 25), which is the lowest proportion among all groups. However, for those who stayed we can observe the highest effort put into further increasing the quantity of output. At the same time this is the only group which significantly improves accuracy of their transcripts. In other words those employees from Group 2 who stayed in the additional time did exert a lot of effort in the additional time. It is interesting to note that participants of this group paid specific attention to the explicit recommendations regarding the quality of output provided in the last set of the instructions and significantly reduced amount of misspellings in the additional time.

The interpretations of the experimental results also found support in the course of the follow-up study. Results of the individual evaluation of the transcripts by the participants of the follow-up study

showed that transcripts of Group 2 enjoyed the highest average scores for perceived quality of preciseness and accuracy. It is interesting that in randomised comparison with transcripts from other groups the transcripts of Group 2 had the lowest chances to be classified as the worst and 13 out of 25 transcripts were chosen in such comparisons as best by at least one evaluator.

Conclusion

Summing up we can say that providing information about the customer in pro-social task context puts the employee into the customer's shoes and allows to move employee's focus towards recognition of the customer's needs. Perceived scope of responsibility for completion of the task and explicit clues regarding the specific needs of the customer help the employee take the decision regarding the individual trade-off between different attributes of output.

Presence of the real individual customer triggers the feeling of responsibility and desire to fulfil customer's need especially among those who believe to be among a few willing to help. In other words presence of the "real customer" who is seen as not helped by others reduces "bystander effect" and motivates the effort.

Managerial perspective

From the managerial perspective the results speak for the advantages of using specific personal information about the individual customer, at least in cases of incomplete contracting where employee works without direct monitoring and is expected to independently take the decisions about quantity / quality trade-off in the interests of the customer. At the same time under such circumstances it is important to provide employees with clear information about their role in completion of the task (single responsibility for the whole task versus part of the task). In case when the employee's perceived responsibility for the task differs from the actual one, she can make a trade-off less beneficial for the customer. Providing clues about the importance of particular output parameters in case of multiattribute tasks can further influence the amount of effort, as well as alter its focus.

One can also expect that employees having stronger personal predisposition for the empathy-related perspective-taking will perform better (more customer-oriented) under such conditions both in terms of general effort and its distribution among different task attributes.

However, we believe that especially for this type of employees this type of interventions requires caution, as "when the weight of the world is on our shoulders, we place ourselves at the risk of burn out" (Grant, 2014).

6.1.4. Exposing the employee to the customer: Group 3

While providing the employee with the information about the customer leads to customer-focused behavior, exposing employee to the customer moves focus back to the self. According to our hypothesis in the context of socially-valued task the requirement to open employee's identity leads to the focus on the attributes of the task easily signalling employee's effort.

As discussed in the Literature review, according to the experimental economics perspective employee should behave more prosocially, or in our context provide higher effort, if her identity becomes open. We make a step further to argue that seemingly prosocial behavior having self-focused motives can be in fact far from benefiting others. Using the multiattribute task we show that seemingly prosocially driven increase in effort can differ from genuinely other-oriented effort in its focus.

According to the agency theory under conditions of direct monitoring in the multiattribute tasks the employee will be interested in providing the higher results on those parameters of output, which can be easier monitored (Holmstrom and Milgrom, 1987).

In our experimental settings our employees are not directly monitored and there are no monetary consequences for weak performance. However, the introduction of the requirement to open the employee's identity can be seen by her as a way of social monitoring. Necessity to provide the name at the end of the transcript made the employee aware that her image could suffer if the customer observed low effort.

In order to show high effort, or in other words, to make sure that the effort is easily recognizable, the employee needs to concentrate on the output parameter easily observable by the customer. While preciseness of the transcript is, for example, the parameter that the customer cannot easily verify, the accuracy of the transcript (amount of mistakes) can be evaluated rather easily. Hence, it is not surprising that in terms of our task accuracy becomes a focus-attribute for Group 3.

Indeed, in the third treatment the general context of the socially desirable task was combined with the requirement to open the identity to the customer, which was in that case the Mitschriften Börse.

The most important result of such manipulation was the clear focus of participants on reducing the number of mistakes. Although the majority of participants, irrespective of the treatment, improved the accuracy of their transcripts in the main time (compared to the control time), in Group 3 this improvement was not only significant, but also the largest among treatments. As we argued in the beginning, we see the accuracy up to a certain point as an important parameter of quality of output. However, taking into account the necessity of balancing different parameters within a limited time-frame, exclusive focus on minimizing mistakes can be also seen as going counter customer's preferences. Once again, we believe that accuracy is the parameter most "easily" characterising output in terms of effort: the misspellings can be seen immediately in the text, while evaluating preciseness requires listening to the original recording. Thus, in order to show that a lot of effort has been put into performing of the task one could try to minimise the visible imperfections of the transcript.

If we compare the behavior of the participants in Group 3 and Group 1 the requirement of opening employee's identity to the customer (but still without direct monitoring by the employer) led to the completely different focus: in particular, the need to follow the requirements of the employer lost its importance. As a result the transcripts provided by the employees of Group 3 have the lowest average level of preciseness among groups, and it is significantly lower than in Group 1. We interpret this result as supporting our hypothesis that introducing the requirement of non-anonymity changes the decision frame and moves attention to self-image.

At the same time the employees in this group tried to maximise the amount of recording transcribed (this was the only group with amount transcribed not significantly different from Group 2). However, this result has been achieved by reducing the importance of preciseness, which is why the average output of this group is characterised by the least amount of words typed among all groups.

Thus, we can conclude that the employees, when required to open their identity, focus on improving the visible and easily verifiable parameters of output (in our case misspellings and mistakes) to the point when they are ready to sacrifice the other parameters more important in terms of satisfying customer needs (in our case preciseness and speed of typing).

Interestingly the individual differences in the level of trait-like empathy (reflecting in our case mostly perspective-taking) had a significant influence on both general level of effort and its focus. In particular those participants who scored higher on the empathic trait typed on average more words and had higher chances of having more precise transcripts. If we take into account the fact that they were not significantly worse in any other output parameter, we can conclude that in this treatment group more empathic employees worked harder and had more balanced approach towards quantity / quality trade-off than their less empathic colleagues.

Supportive of such conclusion is also the fact that the participants of this group had very high chances to stay in the additional time (19 out of 24 employees of this group stayed). Those who stayed also had significantly higher level of agreement with prescriptive norm of helping, than those who did not stay. Thus, we can argue that staying in the additional time was considered as an act of following social norm or in other words image-related decision. At the same time on average the participants of this group did not exert much effort in the additional time. Although they increased the amount of words typed, this increase is only about a half of the increase in Group 2. The focus of effort was directed at improving the formatting of the text: in this parameter Group 3 performed better than any other group in the additional time. We believe that this renders support to our previous conclusion as improving formatting of the transcript (paragraphing, numbering, etc.) can be seen as another "highly visible" parameter of effort.

As a result in the follow-up study we face the situation where the transcripts of this group enjoy the highest evaluation scores in terms of formatting, which means that customers can easily see that the formatting in their transcripts is very good. Interestingly in terms of accuracy the scores for this group have lower mean than for Group 2 and also characterised by larger spread. On average Group 3 has

the smallest amount of high quality transcripts (high scores on parameter "General quality"). This is also the group where 13 out of 24 transcripts were never chosen as best among four transcripts from different groups.

Conclusion

Thus, by exposing the employee to the ("needy") customer we create the frame where the employee is willing to provide customer with what she needs on the one hand, but even stronger she wants to make her effort "visible". In case of low individual empathic predisposition this effort leads to the output unable to meet customer's needs: visible parameters of quality are maximised, while less measurable quality parameters are neglected.

We believe this pattern of behavior is the result of image motivation related to employee's identification by the customer. While opening the identity makes the employee on one hand closer to the customer, on the other it puts her image in question. That is where the desire to transcribe more on one hand and make the transcript visually attractive on the other comes from. The core motive is to *signal* effort.

Managerial perspective

Opening employee's identity if employee performs socially desirable task (e.g. for a needy customer) directs employee's focus towards the attributes of helping behavior signalling the effort to the customer. If the employees score low on the trait-like empathy they fail to put themselves into customer's shoes and as a result the output they provide is unable to meet customer's needs. Thus, self-image focus stimulates the focus directed towards showing the effort, but not necessarily benefiting customer.

6.1.5. Combining effects: Group 4.

Having the image-related hypothesis supported, we move to the results of the last treatment where we combined the conditions introduced in Group 2 and Group 3. Here the participants were provided with information about individual customer and also were required to open their identity to the customer.

As expected the behaviour of the employees also showed the combination of influences, with some more significant effects coming from the effect of non-anonymity requirement. Comparing the results in the main time and control time we see that this was the only group where on average employees did not change significantly their focus. The only significant change was towards improving accuracy: as in Group 3 they decreased the amount of mistakes, but the absolute value of this change is not as dramatic as in Group 3. At the same time unlike in Group 3 this change is less uniform and it is related to the individual differences in the task proficiency. This pattern is similar to the one observed in Group 2. We interpret this result as meaning that requirement to open identity drives the reduction in mistakes, but presence of the information about the customer puts it in a more balanced perspective

where those less proficient do not reduce mistakes dramatically and hence, do not sacrifice other parameters of output.

This interpretation is also supported by the results for the differences in the preciseness which is - similar to Group 1 - positively related to the German proficiency, and - similar to Group 2 - to the level of trait-like empathy.

At the same time looking at the data we observe a large spread between individual results in most of the parameters for this group. Put together, these observations bring us to the conclusion that the participants of Group 4 had a balanced approach while determining the trade-off between different parameters of quantity and quality of output. The information about selected customer helped them to keep the real needs of the customer in mind, while the requirement to open their identity drew attention towards improving the visibility of effort. As a result while maximising quantity those participants who were more proficient could achieve higher quality of output. Still the focus on accuracy was rather strong, with more empathic employees trying to compensate it by extra effort towards preciseness.

Interestingly Group 4 is the one with the highest proportion of participants who stayed in the additional time (21 out 24) and these employees were the only ones significantly improving both of the quantity and quality of output (although in absolute terms improvement in terms of quantity in Group 2 was higher). This group also significantly improved the formatting of their transcripts, which is similar to the pattern observed in Group 3.

Large spread in the individual values of output parameters observed in the main study was also reflected in the evaluations of the transcripts in the follow-up study: this group did not score on average the best in any of the quality parameters, but in all of them they had the widest interquartile range. The general quality scores for the transcripts in this group make it the best among four groups in terms of high-quality transcripts and show that this group had the lowest amount of medium-quality transcripts. Their transcripts had the highest chances to be chosen as the best consequently by two and more evaluators.

Conclusion

By combining the effects of the customer and employee exposure employer can balance the negative influences of image concerns without significantly losing its advantages. While the influence of image-related concerns remains quite strong (the participants significantly reduced the amount of mistakes in the main time and improve formatting in the additional time), individual differences in task proficiency determine the difference in output. While in Group 3 the reduction of mistakes is performed independent of task proficiency, in Group 4 it positively correlates with the participants' speed of typing.

Managerial perspective

If employees are proficient in the tasks they perform, then combining the effects of customer and employee exposure can motivate employees to exert high effort and provide the results best suited for meeting customer needs. However, if there are doubts about employee's proficiency, keeping the condition of employee's anonymity could be a better solution, as it puts less responsibility on employee's shoulders and let her focus on customer's needs and not on self-image.

6.1.6. Limitations of the research

The conclusions discussed above are based on the results of observed differences in the behavior produced in the experimental settings. The applicability of these results significantly depends on the ability of particular experimental design to find an optimal trade-off between external and internal validity, in other words levels of realness of experimental settings and strength of experimental control.

In general, experimental method is characterized by relatively low external validity but high internal validity due to controlled confounds for specific manipulations (Friedman and Cassar, 2004). Low external validity can be reflected in the fact that the behavior of a phenomena observed in the experimental setting are hardly generalizable to the wider real world. The reason for that is that the environments explored in experiments are usually simpler, than complex real environments (Roth, 1995). This can lead to the situation when particular effects important in the experimental settings lose their importance in the real world environment, while other aspects having high importance in the real world have no opportunity to emerge in the experimental setting.

However, a loss of external validity in experimental settings is believed to be balanced by the opportunity to achieve high internal validity. Since internal validity can be influenced by the experimenter, it poses a challenge to eliminate biases and confounds which can compromise it (Antonides, 1991).

Due to the specifics of the research questions set in this thesis, we aimed at creating an experimental design having comparatively higher external validity, while keeping appropriate internal validity. Below we sum up the challenges set by our experimental design as well as the ways these challenges have been answered in present research or how they can be addressed in future research.

Argument 1. The real population differs from the one in the lab and conclusions cannot be applied.

Student's pool. As in most of the laboratory experiments, our subjects pool included exclusively students, thus making the age and work experience distribution different from the general population of interest. However, on average our subjects were 23 years old and had almost two years of work experience, which makes them if not representative of the whole working population, but at least of its junior part.

Self-selection bias. In order to apply the conclusions of the experimental study regarding prosocial behavior to the population outside the lab one should be sure that the appropriate level of randomization of the subject pool. In particular, if only prosocial (or altruistic) subjects self-selected for participation in the experiment, then the results cannot be attributed solely to the treatment effects, but more to their combination with the personality trait representing altruism or prosociality. In order to avoid such self-selection bias we exerted extra care while designing the announcements for participating in the study. In particular we did not open the content of the task and its beneficiary and used monetary incentive as key motivator. We also made sure that the amount paid to the subject was comparative with other similar studies in order to serve as a proper motivation tool.

Argument 2. Specific field of applicability

One-time relationship. In real life working relationships are usually a reoccurring process, which includes the opportunities of reputation building, team work, etc. In our experimental set up employees were performing a one time job. Such design allowed us on one hand control for the reputation effect, but on the other limits the application of the experimental results. It is the goal of the future studies to test the relationship between reputation effect and treatment effects introduced in the present research.

Work versus volunteering. Our experimental design implies that the participants work for a mission-driven organisation of a non-profit type. Undoubtedly these features of design are essentially important for the discussion of applicability of the experimental results. However, we made sure that due to our selection procedure and the special design features employees perceived the task as job rather than volunteering. In particular we attracted and motivated our participants with remuneration comparable to other experiments; we did not explicitly state that the Mitschriften Börse is a non-profit project with volunteers as major workforce. Thus, all our participants work with social purpose in mind, but for real money.

Specifics of the customer. In our experimental setting in order to test the empathy-related hypothesis we used the information about the male customer with disabilities. From the research on empathy, both in psychology and neuroeconomics, we know that empathy can differ for in- and out- groups, race, gender, etc. Further research in necessary to study the links between these characteristics of the employee-customer relationship and their influence on the employee's motivation.

Argument 3. Behavioral change is a result of effects other than treatments

Reputation bias. The results of the experiment could have been influenced by the participants' believes regarding the link between their behavior in the experiment and their grades, etc. In order to avoid this bias we assured that the experimenters were not familiar with subjects. The experimenter though present at the experimental session was there only for answering the questions and helping in case of technical problems. This was also highlighted in the instructions. No monitoring was involved. All data were treated confidentially, which was also explicitly stated in the instructions.

Realness of the task. Participants could be sceptical about the realness of the task and customer. We attempted to make it clear from the instructions that the experiment had real background. Following the logic used in economic experiments (rather than in psychological experiments), we had to motivate our participants by monetary payments and we had to avoid deceiving our participants. That was the reason why we contacted Mitschriften Börse and approached real customer. Question about willingness to finish the task and further participate in the activities of the Mitschriften Börse were also used to reduce the doubts.

Differences in perception of customer needs. Participants could have different understanding of what composed a good transcript. We believe that this problem was significantly reduced by randomization of the participants. This is also supported by the results of the follow-up study, which showed that participants were rather similar in their interpretation of which output parameters were important, and that the differences in the composition of final output could be attributed to the treatment effects.

Differences in perceived responsibility. The results in Group 2 showed that the beliefs about helping behavior of others could significantly influence the trade-off points between task attributes. We interpreted this result in terms of the perceived responsibility for the output, with the perception of single responsibility moving the focus towards the output parameters mostly important for the customer. In our experimental design we did not explicitly state whether the participants were singularly responsible for the amount of output provided. Instead we set a maximisation task. In light of the above results we believe it would be very interesting to provide an explicit test of the influence the perceived responsibility has on the attributes trade-offs. For that in future research we suggest introducing the fixed quantitative goal and explicit information regarding the responsibility for the task.

Argument 4. Individual differences in cost of effort (intrinsic motivation, task proficiency) could influence the results beyond level described

Was it empathy? In order to measure the level of empathy (both dispositional and situational) we used the self-reported measures, which means that they can suffer from desirability bias and other biases related to this approach. At the same time the instruments used were either extensively simplified and shortened versions (for trait-like empathy), or versions not fully adjusted for the current subject pool (for state-like empathy, where we used the list of adjectives in English language unfortunately without providing their translation or explanation in German). The reason, as was explained in the Experimental design section, was to reduce the time and cognitive load represented by these questionnaires on participants as well as to avoid priming them in a specific direction. In order to get a better picture of interrelation between personal traits, states and experimental effects we believe future research should still move the trade-off point towards the deeper focus on self-reported measures, thus implying using more detailed measures of dispositional and situational empathy.

Level of task proficiency. Differences both in the overall amount of output and in the trade-offs between different attributes of output could be a result of variation in task specific skills between participants. Beyond the standard experimental tool of random assignment to the groups, which reduces the influence of individual differences on group means, we introduced several other safeguards in this respect. In particular, the requirements for the participants set in the announcements included levels of language skills, both German and English, as well as excellent typing skills. Within the experimental procedure we took account of task-specific skills by including a training time, when each participant could try to transcribe a recording, using the provided interface and have her questions answered. In the statistical analysis the individual results in the control time served as the explicit mechanisms for controlling for the differences in task-specific skills and other individual differences (age, gender, experience, etc.). In future research in order to explicitly test the relationship between the differences in skills and treatment effect more precise measurement of skills can be introduced, in particular by stronger focus on related measures design (using longer control time for measuring "pure" effort).

Omitted variables. There are many other variables – especially in psychological stream of research – which could potentially be influencing the results of interactions similar to the one highlighted in our experimental design. To name just one, Graso and Probst (2012) discuss the influence of individual differences in consideration of future consequences (CFC) on quantity-quality trade-offs in multiattribute tasks.

Researchers address individual differences as reasons for differences in the performance. Their focus is consideration of future consequences, which relates to the notion of individual's temporary orientation. CFC relates to the individual consideration of short- and long-term outcomes (introduced by Strathman et al, 1994), which has been shown to predict environmental and health behavior better than conscientiousness and other temporal orientation instruments

According to Graso and Probst (2012) employee's performance at work is generally dependent on the type of work performed, incentives utilized and congruence between those factors and the personal characteristics of the employee (knowledge, skills, abilities, and others). Authors suggest that employee's performance may be predicted by congruence between CFC orientation and quantity versus. quality demands of the work itself.

In their experiment people were performing a data-entry task with the requirement to maximize the amount typed and minimize the number of errors. People low on CFC focused more on quantitative parameters of the task, and people with high CFC – on qualitative. They argue that CFC orientation may unconsciously drive which aspects of performance are seen as most salient and, therefore needing attention.

We believe that future experimental research is necessary to validate the results of experimental study and open new aspects of relationships between identification and motivation in employee-customer dyad.

6.1.7. Main implications of the research

Summing up the discussion about the results of present research we can point at several directions in which it attempts to cover the scientific gap currently existing in theory and practice.

First of all we combine the arguments regarding pro-social behavior brought by the research in the field of experimental economics with the context-dependent approach of management and organizational psychology and sociology and apply this knowledge to provide new insights into the field of job motivation.

We widen the frame studying monetary or "pure" incentives in terms of Andreoni and Rao (2011) common in economic experiments to include the social component of giving. In particular, we incentivise the participation by purely monetary incentives, but let the effort be motivated by the customer-related concerns.

Such interdisciplinary approach in its turn allows us to relate the concept of other-regarding behavior to the patterns of behavior within customer-employee dyad. Moreover in the present research other-regarding behaviour towards customer is put into the centre of discussion about job motivation.

Without questioning the importance of classical motivational mechanisms ranging from pure monetary incentives, career concerns in economics or intrinsic motives in management, we focus on the specific psychological needs of employees – namely, meaningfulness of their jobs – and see the opportunity to benefit the customer as a way to satisfy such need.

Further we accept the context-dependent view of the pro-social behavior, and, importantly, we move on to search for its proximal motives. In doing this we agree with the arguments of the researchers in the field of identity economics who believe that social context can have dramatic effect on decisions.[83]

While pecuniary motivations or preferences – the desire to consume goods and services – are privileged in traditional economic theories and seen as idiosyncratic, static and context-free, social contexts tend to dominate many aspects of economic lives (Huettel and Kranton, 2012). Andreoni and Rao (2011) argue that many theories trying to explain prosocial behavior rely on the fact that humans are social species. These theories are based on the mechanisms of social interaction (e.g. communication for reputation building) and social cues (which allow people to be not indiscriminately altruistic).

In our research the participants take the decisions (make trade-offs), because they have a mixture of preferences, and the environment (context) triggers the change in importance of some of these preferences. While all people have some mixture of preferences (for different attributes of outcome),

---

[83]Identity economics tries to identify non-pecuniary sources of motivation and the relations between these motivation and social context. It focuses, at its core, on individuals who make decisions in a social context, bringing identity and social context into utility function (Huettel and Kranton, 2012).

the importance of each attribute can differ depending on the contextual clues. This means that we can see the fact that people have the other-regarding preferences as a trait (context-independent), but the weights people assign to such preferences (if they do) as context dependent.

As a result of both interdisciplinary literature review and our own experimental manipulations we come to the conclusion that other-regarding behaviour is a complex construct having other-benefiting and self-benefiting motives. Furthermore these motives have specific driving forces standing behind them. We distinguish between two of such forces, namely empathy and image concerns.

We apply the context-dependent frame to the discussion of empathy and image-concerns and show that customer and employee identification can help to create the contexts allowing for empathy and image concerns to govern employee's behaviour towards customer and influence the level of employee's effort and parameters of her output.

In order to be able to test the theoretical hypotheses it was important to select a research method allowing for maximal closeness to the real job environment, at the same time allowing sufficient level of external control.

As Charness and Gneezy (2000) admit, there are rather few studies in economics that explore the influence of social context on behavior. As it is standard in economics experiments to maintain anonymity among the participants, the attempt is rather to increase the social distance, which is done due to concerns for the potential loss of control in experimental settings. Yet Charness and Gneezy (2000) point out that such concerns can be most appropriate for testing a theory of pure selfishness, but if we assume that people are not entirely selfish, the interest in the patterns of non-selfish concerns justifies different approaches to experimental testing.

While our choice favoured the experimental method, in design of the experiment, as in the theoretical analysis, we used interdisciplinary approach. In particular

- While most of the economic experiments use money as the currency of the experiment, we used individual effort of the participants. In other words participants' decisions were influencing not the amount of monetary remuneration, but the amount of effort they put into performing the task. It is also very important that the participants had to take decisions not about imaginary amount of effort, but had to perform the real task and exert effort on the spot.

- Whereas in some economic experiments participants perform real tasks, they are not performed for the real existing customer and thus, lack meaning. Our experimental design allowed us to introduce the employees to the task, where their effort was benefiting the real customer making the task-related activity meaningful.

- There are experiments conducted in the field of organizational psychology and management which include real and meaningful tasks, but subjects are usually provided with untruthful information and rarely paid with real money. In order to avoid deceiving our subjects in our experimental design we found a setting with the real customer and we paid our employees a real monetary fee.

At the same time we accompanied the remunerated part by the opportunity to exert effort on voluntary terms in order to see, whether the behavior of the same employee changes in the absence of monetary reward.

Most importantly, this is one of the very few experimental studies in the field of management where participants had to exert effort in the multiattribute task (and to the best of our knowledge, it is the only study using four different output dimensions).

Multitasking environments have been extensively researched in economics and the general prescriptions were related to usage of high versus low-powered incentives for balancing the distribution of effort among the task. For example Burgess and Ratto (2003) argue that if actions are substitutes, high-powered incentives may have undesirable effect on performance: higher marginal incentives in one task drive the agent's efforts from the substitute tasks. So in a context with multiple dimensions of output this makes agent concentrate on the output which is more accurately measured. Or in some cases agent can concentrate on the easier tasks (if for them they are substitutes, but compliments for the principle).

In our analysis we use fixed payment scheme which does not provide any explicit incentives for particular tasks. At the same time we believe that non-monetary tools can be used to move the employees towards optimal trade-offs in multiattribute tasks. In particular we use different types of social information as motivators and test their effect on the focus of effort.

In order to make sure that the results of our experiment can be attributed to the differences in treatments we used the combination of related samples and independent samples designs. Furthermore we follow Friedman and Cassar (2004) in their statement that laboratory experiments are no substitute for the conduct and the analysis of questionnaire data, and complemented our experimental design with questionnaires.

Another feature of our research constitutes in supplementing the results of behavioral experiment by the results of the follow-up study, aiming at discovering of common beliefs about the effects of the treatment interventions on the quality/quantity trade-offs in the multiattribute task. While the specifics of multiattribute task allows us to see the difference in the outcomes of the decision process, in order to make conclusions we ought to be sure that the link between specific task attributes and treatment effect is seen by the participants the same way as hypothesised by the experimenters. We believe that the strength of such link can be tested by means of eliciting the beliefs of the participants themselves about the importance of different output parameters in different treatment conditions. It is important that the participants are motivated to take the beliefs elicitation process seriously and exert effort to get as close as possible to the true values (values one could observe in the population).

We conduct such game within the follow-up study where we ask participants to evaluate the relative importance of different parameters of output for the participants from each of four treatments. To

motivate this process we include a winning option for providing evaluation closest to the results of our experimental study (of course unknown to the participants).

Another new feature of our design was related to the complexity of the output structure, which implied that the maximization of single parameters was not necessarily leading to the increase in customer satisfaction. The decisions regarding the trade-offs among individual parameters of output required the ability and desire to see the customer satisfaction as the final goal of the effort. That is why extensive focus on one or two output parameters while disregarding the others could result in the output unable to satisfy customer's needs. In order to see whether our interpretation of the main results reflected the opinion of customers we conducted an independent objective evaluation of output within the follow-up study.

All above features of our experimental and post experimental designs were aimed at creating the environment best suited for testing the behavioral hypothesis in the job context, as well as interpreting the results in most objective and realistic way. The results should be seen as a missing piece of puzzle, though small, but still allowing to connect several streams of research by focusing the efforts on understanding the forces and mechanisms behind the complex construct of work motivation.

List of References

Ackert, L. F., Church, B. K., and Davis, S. (2006) Social Distance and Reciprocity. Working paper.

Akerlof, G. A. (2007) The Missing Motivation in Macroeconomics. *American Economic Review, 97,* 1, 5-36.

Alkire, S., and Séverine, D. (2000) Individual Motivation, Its Nature, Determinants and Consequences for Within Group Behaviour. Helsinki: UNU World Institute for Development Economics Research.

Andreoni, J. (1990) Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving. *The Economic Journal, 100(401)*, 464-477.

Andreoni, J. (1998) Toward a Theory of Charitable Fund-Raising. *Journal of Political Economy, 106(6)*, 1186-1197.

Andreoni, J., and Miller, J. (2002) Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism. *Econometrica Evanston Ill-, 70,* 737-754.

Andreoni, J., and Rao, J. M. (2011) The Power of Asking: How Communication Affects Selfishness, Empathy, and Altruism. *Journal of Public Economics, 95,* 513-520.

Andreoni, J., Rao, J., and Trachtman, H. (2011) Avoiding the Ask A Field Experiment on Altruism, Empathy, and Charitable Giving. Cambridge, Mass: National Bureau of Economic Research.

Antonides, G. (1991) Psychology in Economics and Business. Dordrecht: Kluwer Academics.

Ariely, D., Bracha, A., and Meier, S. (2009) Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially. *American Economic Review, 99(1)*, 544-555.

Ariely, D., Prelec, D., and Kamenica, E. (2008) Man's Search for Meaning: The Case of Legos. *Journal of Economic Behavior and Organization, 67*, 671-677.

Ariely, D. (2005) Tom Sawyer and the Construction of Value. Federal Reserve Bank of Boston Boston, Mass.

Baron-Cohen, S., Richler, J., Bisarya, D., Gurunathan, N., and Wheelwright, S. (2003) The Systemizing Quotient: an Investigation of Adults with Asperger Syndrome or High-Functioning Autism, and Normal Sex Differences. *Philosophical Transactions: Biological Sciences, 358*, 1430, 361-374.

Bartlett, M.Y, and D DeSteno. 2006. "Gratitude and Prosocial Behavior: Helping When It Costs You". *Psychological Science.* 17, no. 4: 319-25.

Batson, C. D., Cochran, P.J., Biederman, M. F., Blosser, J.L., Ryan, M.J. and Vogt, B. (1978) Failure to Help When in a Hurry: Callousness or Conflict? *Personality and Social Psychology Bulletin,* 4 (1): 97-101.

Batson C.D., Eklund J.H., Chermok, V.L., Hoyt, J.L., and Ortiz, B.G. (2007) An Additional Antecedent of Empathic Concern: Valuing the Welfare of the Person in Need. *Journal of Personality and Social Psychology.* 93 (1): 65-74.

Batson, C. D., and Dani C. (1983) Influence of Self-Reported Distress and Empathy on Egoistic Versus Altruistic Motivation to Help. *Journal of Personality and Social Psychology.* 45 (3): 706-718.

Batson, C.D. (1984) *Distress and Empathy: Two Qualitatively Distinct Vicarious Emotions with Different Motivational Consequences*. Advances in Experimental Social Psychology, 20: 65-120.

Batson, C.D. (1987). Prosocial Motivation: Is it ever Truly Altruistic? In: L. Berkowitz, (Ed.), New York: Academic Press.

Batson, C.D. (2008) Empathy-Induced Altruistic Motivation. Draft of lecture/chapter for Inaugural Herzliya Symposium on "Prosocial Motives, Emotions, and Behavior," March 24-27.

Bénabou, R. and Tirole, J. (2006) Incentives and Prosocial Behavior. *American Economic Review*, 96(5): 1652-1678.

Ben-Ner, A. (2013) Preferences and Organization Structure: Toward Behavioral Economics Micro-Foundations of Organizational Analysis. *The Journal of Socio-Economics.* 46 (6): 87-96.

Besley, T. and Ghatak, M. (2005) Competition and Incentives with Motivated Agents. *The American Economic Review*, 95 (3): 616-636.

Bicchieri, C. and Chavez, A. (2010) Behaving as Expected: Public Information and Fairness Norms. *Journal of Behavioral Decision Making*, 178 (2): 161-178.

Bicchieri, C. (2006) The grammar of society: the nature and dynamics of social norms. New York: Cambridge University Press.

Bohnet, I. and Frey, B. S. (1999). Social Distance and Other-Regarding Behavior in Dictator Games. *American Economic Review, 89,* 1: 335-339.

Bolton, G. E. and Ockenfels, A. (2000) ERC: A Theory of Equity, Reciprocity, and Competition. *American Economic Review*, 90(1): 166–193.

Brandts, J. and Charness, G. (2011) The Strategy Versus the Direct-Response Method: a First Survey of Experimental Comparisons. *Experimental Economics.* 14 (3): 375-398.

Brüggen, A. and Strobel, M. (2007) Real effort versus chosen effort in experiments. *Economics Letters*, 96: 232–236.

Brülhart, M. and Usunier, J.C.. (2008) Verified Trust: Reciprocity, Altruism, and Randomness in Trust Games. Lausanne: IRM.

Burgess, S. and Ratto, M. (2003) The Role of Incentives in the Public Sector: Issues and Evidence. *Oxford Review of Economic Policy.* 19 (2): 285-300.

Calabrese, A. (2012) Service Productivity and Service Quality: A Necessary Trade-Off?. *International Journal of Production Economics,* 135 (2): 800-812.

Camerer, C. F. (2003). Behavioral Game Theory: Experiments in Strategic Interaction. Princeton, NJ: Princeton Univ. Press.

Chalofsky N. and Krishna, V. (2009) Meaningfulness, Commitment, and Engagement:The Intersection of a Deeper Level of Intrinsic Motivation. *Advances in Developing Human Resources,* 11 (2): 189-203.

Chalofsky, N. (2003). An emerging construct for meaningful work. *Human Resource Development International, 6,* 1: 69-83.

Charness, G., and Rabin, M. (2002). Understanding social preferences with simple tests. *The Quarterly Journal of Economics, 117,* 3: 817-869.

Charness, G. and Gneezy, U. (2000). What's in a Name? Anonymity and Social Distance in Dictator and Ultimatum Games. eScholarship, University of California.

Charness, G., and Schram, A. (2012) *Social and Moral Norms in the Laboratory*. UC Santa Barbara: Department of Economics, UCSB.

Christensen P. N., Rothgerber, H., Wood, W. and Matz, D. C. (2004) Social Norms and Identity Relevance: A Motivational Approach to Normative Behavior. *Personality and Social Psychology Bulletin.* 30 (10): 1295-1309.

Cialdini, R. B. and Goldstein, N. J. (2004) Social Influence: Compliance and Conformity. *Annual Review of Psychology, 55 (*1), 591-621.

Cialdini, R. B. (1991) Altruism or Egoism? That Is (Still) the Question". *Psychological Inquiry.* 2 (2): 124-126.

Coke, J. S., Batson C.D. and MacDavis, K. (1978) Empathic Mediation of Helping: A Two-Stage Model. *Journal of Personality and Social Psychology.* 36 (7): 752-766.

Cryder, C. and Loewenstein, G. The critical link between tangibility and generosity. Published in Oppenheimer, D. M., Olivola, C. Y., and Society for Judgment and Decision Making. (2011). *The science of giving: Experimental approaches to the study of charity*. New York: Psychology Press. (pp.237-252).

Delfgaauw, J. and Dur, R. (2008). Incentives and Workersandrsquo; Motivation in the Public Sector. *The Economic Journal, 118,* 525: 171-191.

Delfgaauw, J. and Dur, R.. (2007) Incentives and Workers' Motivation in the Public Sector*. *The Economic Journal.* 118 (525): 171-191.

Dickert, S., Sagara, N., and Slovic, P. Affective motivation to help others: a two-stage model of donation decisions Published in Oppenheimer, D. M., Olivola, C. Y., and Society for Judgment and Decision Making. (2011). *The science of giving: Experimental approaches to the study of charity*. New York: Psychology Press.(pp. 161-178)

Van Dijk, F., Sonnemans, J. and van Winden, F. (2001). Incentive Systems in a Real Effort Experiment. *European Economic Review.* 45 (2): 187-214.

Dixit, A., 2002. Incentives and Organizations in the Public Sector: An Interpretative Review. *The Journal of Human Resources*, 37(4), 696-714.

Dreber, A., Ellingsen T., Johannesson, M. and Rand, D.G. (2011). Do people care about context?: framing effects in dictator games. Stockholm: Ekonomiska Forskningsinst.

Edele, A., Dziobek, I. and Keller, M. (2013) Explaining Altruistic Sharing in the Dictator Game: The Role of Affective Empathy, Cognitive Empathy, and Justice Sensitivity. *Learning and Individual Differences.* 24 (3): 96-102.

Einolf, C.J. (2008) Empathic Concern and Prosocial Behaviors: A Test of Experimental Results Using Survey Data. *Social Science Research,* 37 (4): 1267-1279.

Ellingsen, T. and Johannesson, M. (2006) Pride and Prejudice: The Human Side of Incentive Theory. London: Centre for Economic Policy Research.

Ellingsen, T., Johannesson, M., Mollerstrom, J. and Munkhammar, S. (2012) Social Framing Effects: Preferences or Beliefs? *Games and Economic Behavior,* 76 (1): 117-130.

Fahr, R. and Irlenbusch, B. (2000) Fairness As a Constraint on Trust in Reciprocity: Earned Property Rights in a Reciprocal Exchange Experiment. *Economics Letters,* 66 (3): 275-282.

Faillo, M. and Sacconi, L. (2007) Norm compliance: the contribution of behavioral economics models. Discussion Paper No.4.

Fehr, E. and Fischbacher, U. (2002) Why Social Preferences Matter - the Impact of Non-Selfish Motives on Competition, Cooperation and Incentives. *The Economic Journal*, 112 (478): 1-33.

Fehr, E. and Schmidt, K.M. (1999) A Theory Of Fairness, Competition, and Cooperation. *Quarterly Journal of Economics*, 114(3): 817-868.

Fehr, E. and Fischbacher, U. (2005). Human Altruism: Proximate Patterns and Evolutionary Origins. *Analyse and Kritik*, 27 (1): 6-16.

Fehr, E. and Singer, T. (2005) *The neuroconomics! of mind reading and empathy*. London: Centre for Economic Policy Research.

Fehrler, S. and Kosfeld M. (2014) Pro-Social Missions and Worker Motivation: An Experimental Study. *Journal of Economic Behavior and Organization.* 100 (4): 99-110.

Francois, P. (2000) 'Public service motivation' as an argument for government provision. *Journal of Public Economics.* 78 (3): 275-299.

Francois, P. and Vlassopoulos, M. (2008) Pro-social motivation and the delivery of social services. CESifo Economic Studies, 54(1): 22-54.

Francois, P. (2007) Making a difference *RAND Journal of Economics,* 38 (3): 714–732

Frankl, V.E. (2006) Man's Search for Meaning. Boston Beacon Press (1962).

Frey, B.S. and Meier, S. (2004). Pro-Social Behavior in a Natural Setting. *Journal of Economic Behavior and Organization,* 54 (1): 65-88.

Friedman, D., and Cassar, A. (2004). *Economic Lab: An Intensive Course in Experimental Economics.* New York and London: Routledge.

Gächter, S. and Fehr, E. (2002) Fairness in the labour market – A survey of experimental results. Working paper No.114 published in Bolle, F., and Lehmann-Waffenschmidt, M. (2002). Surveys in experimental economics: Bargaining, cooperation and election stock markets. Heidelberg: Physica-Verlag.

Gallucci, M. and Perugini, M. (2000). An experimental test of a game-theoretical model of reciprocity. *Journal of Behavioral Decision Making*, 13: 367-389.

Gehlbach, H., Brinkworth, M.E. and Wang, M. (2012) The Social Perspective Taking Process: What Motivates Individuals to Take Another's Perspective? *Teachers College Record.* 114 (1).

Gintis, H. (2005). Behavioral Game Theory and Contemporary Economic Theory. *Analyse and Kritik, 27 (*1): 48.

Grant, A. M. (2014) The #1 Feature of a Meaningless Job, Internet article posted: 01.30.2014. http://www.huffingtonpost.com/adam-grant/the-1-feature-of-a-meanin_b_4691464.html

Grant, A. M. (2008) The Significance of Task Significance: Job Performance Effects, Relational Mechanisms, and Boundary Conditions. *Journal of Applied Psychology, 93 (*1): 108.

Grant, A. M. and Berg, J. M. (2010) Prosocial motivation at work: How making a difference makes a difference. Forthcoming in K. Cameron and G. Spreitzer (Eds.), Handbook of Positive Organizational Scholarship. Oxford University Press.

Grant, A. M., and Berry, J. W. (2011) The Necessity of Others Is The Mother of Invention: Intrinsic and Prosocial Motivations, Perspective Taking, and Creativity. *Academy of Management Journal.* 54 (1): 73-96.

Grant, A. M., and Parker, S. K. (2009). 7 Redesigning Work Design Theories: The Rise of Relational and Proactive Perspectives. *The Academy of Management Annals, 3 (*1): 317-375.

Grant, A. M. (2007) Relational Job Design and the Motivation to Make a Prosocial Difference. *The Academy of Management Review.* 32 (2): 393-417.

Grant, A. M. (2008) The Significance of Task Significance: Job Performance Effects, Relational Mechanisms, and Boundary Conditions. *Journal of Applied Psychology,* 93 (1): 108-124.

Grant, A. M. and Hofmann, D.A. (2011) Outsourcing Inspiration: The Performance Effects of Ideological Messages from Leaders and Beneficiaries. *Organizational Behavior and Human Decision Processes,* 116 (2): 173-187.

Grant, A. M. (2008) Employees without a Cause: The Motivational Effects of Prosocial Impact in Public Service. *International Public Management Journal,* 11 (1): 48-66.

Graso, M. and Probst, T.M. (2012) The Effect of Consideration of Future Consequences on Quality and Quantity Aspects of Job Performance. *Journal of Applied Social Psychology,* 42 (6): 1335-1352.

Grimalda, G. and Sacconi, L. (2005) The Constitution of the Not-For-Profit Organisation: Reciprocal Conformity to Morality. *Constitutional Political Economy,* 16 (3): 249-276.

Hackman J.R. and Oldham, G.R. (1976) Motivation through the design of work: test of a theory, *Organizational Behavior and Human Performance*, 16 (2): 250-279.

Hamilton, W. D. (1963). The evolution of altruistic behavior. *American Naturalist* 97 (896): 354–356.

Holmstrom, B. and Milgrom, P. (1987) Aggregation and Linearity in the Provision of Intertemporal Incentives. *Econometrica.* 55 (2): 303-328.

Huettel, S. and Kranton, R. (n.d.) Identity Economics and the Brain: Uncovering Mechanisms of Social Conflict. *Philosophical Transactions of the Royal Society B Biological Sciences*.

Johnson, N.D. and Mislin, A.A. (2011) Trust games: A meta-analysis. *Journal of Economic Psychology,* 32 (5): 865-889.

Kahneman, D. (2013) Thinking, Fast and Slow. New York: Farrar, Straus and Giroux.

Kogut, T., and Ritov, I The idenifiable victim effect: causes and boundary conditions Published in Oppenheimer, D. M., Olivola, C. Y., and Society for Judgment and Decision Making. (2011). *The science of giving: Experimental approaches to the study of charity*. New York: Psychology Press.(133-148)

Konow, J. (2010) Mixed Feelings: Theories of and Evidence on Giving. *Journal of Public Economics*, 94 (3-4): 279-297

Kosfeld, M., Neckermann, S. and Yang, X. (2013) Knowing That You Matter, Matters! The Interplay of Meaning, Monetary Incentives, and Worker Recognition. Bonn: IZA.

Leider, S., Mèobius, M.M., Rosenblat, T. and Do, Q. (2009) Directed Altruism and Enforced Reciprocity in Social Networks. *Quarterly Journal of Economics,* 124 (4): 1815-1851.

Leider, S., Mèobius, M.M., Rosenblat, T. and Do, Q. (2010) What Do We Expect from Our Friends? *Journal of the European Economic Association,* 8 (1): 120-138.

Lester, S. W., Meglino, B.M.and Korsgaard, M.A. (2008) The Role of Other Orientation in Organizational Citizenship Behavior. *Journal of Organizational Behavior,* 29 (6): 829-841.

Levitt, S.D. and List, J.A. (2007) Viewpoint: On the Generalizability of Lab Behaviour to the Field. *Canadian Journal of Economics/Revue Canadienne D'économique.* 40 (2): 347-370.

List, J. A. (2007). On the Interpretation of Giving in Dictator Games. *Journal of Political Economy, 115 (*3): 482-493.

Loewen, P.J., Lyle, G. and Nachshen, J.S. (2009) An eight-item form of the Empathy Quotient (EQ) and an application to charitable giving.

Maslow, A. H. (1971) The Farther Reaches of Human Nature, New York: Penguin

Meier, S. (2006) A Survey of Economic Theories and Field Evidence on Pro-Social Behavior. Boston, Mass: Federal Reserve Bank of Boston.

Moynihan, D.P., Pandey, S.K. and Wright, B.E. (2012) Prosocial Values and Performance Management Theory: Linking Perceived Social Impact and Performance Information Use. *Governance.* 25 (3): 463-483.

Murdock, K. (2002) Intrinsic Motivation and Optimal Incentive Contracts. *RAND Journal of Economics.* 33 (4): 0650-0671.

Murphy R.O., Ackermann K.A., and Handgraaf M.J. (2011) Measuring Social Value Orientation. *Judgment and Decision Making.* 6 (8): 771-781.

Oppenheimer, D. M., Olivola, C. Y., and Society for Judgment and Decision Making. (2011). *The science of giving: Experimental approaches to the study of charity*. New York: Psychology Press.

Peacock, M. S., Schefczyk, M., and Schaber, P. (2005). Altruism and the Indispensability of Motives. *Analyse and Kritik, 27 (*1): 188.

Povey, R. (2014) The limits to altruism. Working paper. Internet resource http://users.ox.ac.uk/~sedm1375/Working%20Papers/tlasart.pdf

Prendergast, C. (2007) The Motivation and Bias of Bureaucrats. *American Economic Review*, 97(1): 180-196.

Prendergast, C. (2008) Intrinsic Motivation and Incentives. *American Economic Review*, 98(2): 201-205.

Preston, S. D., and de Waal, F. B. (2002). Empathy: Its ultimate and proximate bases. *Behavioral and Brain Sciences, 2 (*1): 1-20.

Rauhut, H. and Winter, F. (2010). A sociological perspective on measuring social norms by means of strategy method experiments. *Social Science Research, 39 (*6): 1181-1194.

Rege, M., and Telle, K. (2004). The impact of social approval and framing on cooperation in public good situations. *Journal of Public Economics, 88 (*7): 1625.

Roth, A. (1995) Bargaining Experiments. In: Kagel, J.H., Roth, A.E. (eds) Handbook of Experimental Economics. Princeton University Press: Princeton. (253-348).

Schelling, T. C. (1968). The Life You Save May Be Your Own. In Samuel Chase (Ed.) Problems in Public Expenditure Analysis. Washington DC: The Brookings Institute.

Schulz, J.F., Fischbacher, U., Thöni, Ch. and Utikal, V. (2014). Affect and Fairness: Dictator Games Under Cognitive Load. *Journal of Economic Psychology (*41): 77-87.

Simon, H. A. (1991) Organizations and Markets. *The Journal of Economic Perspectives (Nashville),* 5 (2): 25-44.

Singer, T., and Lamm, C. (2009). The Social Neuroscience of Empathy. *Annals of the New York Academy of Sciences*, 1156: 81-96.

Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R.J. and Frith, Ch. D. (2004) Empathy for Pain Involves the Affective but Not Sensory Components of Pain. *Science.* 303 (5661): 1157-1162.

Small, D. Sympathy biases and sympathy appeals: reducing social distance to boost charitable contributions Published in Oppenheimer, D. M., Olivola, C. Y., and Society for Judgment and Decision Making. (2011). *The science of giving: Experimental approaches to the study of charity*. New York: Psychology Press.(149-160)

Small, D. and Loewenstein, G. (2003) Helping a Victim or Helping the Victim: Altruism and Identifiability. *Journal of Risk and Uncertainty.* 26 (1): 5-16.

Smith, A. (1759/1976) The Theory of Moral Sentiments. (Raphael, D.D.; Macfie, A.L., eds.), Oxford: Claredon Press.

Smith, A. (1776/1937). An inquiry into the nature and causes of the wealth of nations. In Cannan, E., & Lerner, M. New York: The Modern library.

Terkel, S. (1974) Working: People Talk About What They Do All Day and How They Feel About What They Do. NY: Pantheon/Random House.

Thakor, M. V. and Joshi, A. W. (2005). Motivating salesperson customer orientation: insights from the job characteristics model. *Journal of Business Research, 58 (*5): 584.

Toi, M. and Batson, C.D. (1982) More Evidence That Empathy Is a Source of Altruistic Motivation. *Journal of Personality and Social Psychology.* 43 (2): 281-292.

Tonin, M. and Vlassopoulos, M. (2009) *Disentangling the Sources of Pro-Social Behavior in the Workplace: A Field Experiment*. Southampton: University of Southampton, School of Social Sciences, Economics Division.

Tonin, M. and Vlassopoulos, M. (2012) *Social Incentives Matter Evidence from an Online Real Effort Experiment*. Bonn: IZA.

Vignemont, de F. and Singer, T. (2006). The empathic brain: how, when and why? *Trends in Cognitive Sciences,* 10: 435-441.

Wakabayashi, A., Baron-Cohen, S., Wheelwright, S., Goldenfeld, N., Delaney, J., Fine, D., Smith, R. and Weil, L. (2006) Development of Short Forms of the Empathy Quotient (EQ-Short) and the Systemizing Quotient (SQ-Short). *Personality and Individual Differences.* 41 (5): 929-940.

Winden van, F., Van Dijk, F. and Sonnemans, J. (2000) Incentive Systems in a Real Effort Experiment. CESifo Working Paper Series, No. 272.

Wright, B.E. (2003). Toward understanding task, mission and public service motivation: A conceptual and empirical synthesis of goal theory and public service motivation. Presented at the 7th National Public Management Research Conference, Washington, D.C.

Wright, B.E. and Grant, A.M. (2010) Unanswered Questions About Public Service Motivation: Designing Research to Address Key Issues of Emergence and Effects. *Public Administration Review.* 70 (5): 691-700.

Appendixes

Dear Students of the Faculty of Business and Economics!

Chair of Organization and Planning of the Faculty of Business and Economics invites you to take part in the study of organizational behavior!

When: **6ᵗʰ and 8ᵗʰ of May, 2013**

Duration: Around **2 hours**

Remuneration: All participants will earn a fixed amount of **20 Euro**, which will be paid in cash at the end of the study

Requirements: **Good English** and **excellent German** knowledge, **good typing** skills

Registration: In order to register for participation in the study, please send an email to **experiment.orga@univie.ac.at** with the following information:

Full name_____

Student ID number_____

Email_____

Mobile phone number_____

*Preferable participation time (*Please indicate **all** options suitable for you!*):*

__6 May, 2013:__ 8.30-11.00 / 11.30-14.00 / 14.30-17.00 / 17.30-20.00

__8 May, 2013:__ 11.30-14.00 / 14.30-17.00 / 17.30-20.00

Registration deadline: We kindly ask you to send your registration request until **28 April, 2013**!

Since the number of participants is limited, early registration will increase your chances!

Thank you for your support!

Instructions for the control time (Part 1)

Instructions

Dear participants!

Thank you for taking part in the study of organizational behavior. This session will last for about 2 hours, during which you will perform a **job task** and fill in a **set of questionnaires**. You are paid a fixed amount of **5 Euro for answering four questionnaires** and a fixed amount of **15 Euro for performing a job task.**

All the information provided in the questionnaires will be kept confidential.

Please keep in mind that you are not allowed to talk during the whole session. If you have any questions, please raise your hand and we will come up to you to answer.

Please start from answering the two questionnairs. Remember that there are no right or wrong answers; the issue of concern is your personal preferences and feelings. Your answers do not influence your remuneration for filling the questionnaires.

You have about 10 minutes for completing this part.

Questionnair_1

Imagine that you have been randomly paired with another person, whom we refer to as **the other**. This other person is someone you do not know and you will remain mutually anonymous. Imagine that you have money, which should be split between you and this other person.

On the next page you will find six choice options. Please deside how you would split the money between you and the other person in each option.

For instance, in the example below a person has chosen such a distribution, where he/she receives 50 Euro, while the anonymous other person receives 40 Euro.
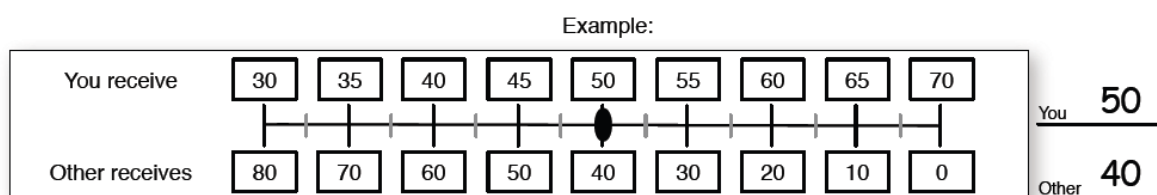
In each of the following six choice options, please mark the prefered distributiion along the line (draw a black dot) and write the resulting distibution of money on the spaces on the right (as in the example above).

Please remember that there are no right or wrong answers, this is all about personal preferences.

1

You receive	85	85	85	85	85	85	85	85	85	You _____
Other receives	85	76	68	59	50	41	33	24	15	Other _____

2

You receive	85	87	89	91	93	94	96	98	100	You _____
Other receives	15	19	24	28	33	37	41	46	50	Other _____

3

You receive	50	54	59	63	68	72	76	81	85	You _____
Other receives	100	98	96	94	93	91	89	87	85	Other _____

4

You receive	50	54	59	63	68	72	76	81	85	You _____
Other receives	100	89	79	68	58	47	36	26	15	Other _____

5

You receive	100	94	88	81	75	69	63	56	50	You _____
Other receives	50	56	63	69	75	81	88	94	100	Other _____

6

You receive	100	98	96	94	93	91	89	87	85	You _____
Other receives	50	54	59	63	68	72	76	81	85	Other _____

Questionnair_2

Please indicate on a 7-point scale (from *not at all* to *very much*), how well each of these statements describe yourself:

	Not at all						**Very much**
I find it easy to put myself in somebody else's shoes	1	2	3	4	5	6	7
I find it hard to know what to do in a social situation.	1	2	3	4	5	6	7
I am good at predicting how someone will feel.	1	2	3	4	5	6	7
I often find it hard to judge if something is rude or polite.	1	2	3	4	5	6	7
I am quick to spot when someone in a group is feeling awkward or uncomfortable.	1	2	3	4	5	6	7
Other people often say that I am insensitive, though I do not always see why.	1	2	3	4	5	6	7
Other people tell me I am good at understanding how they are feeling and what they are thinking.	1	2	3	4	5	6	7
It is hard for me to see why some things upset people so much.	1	2	3	4	5	6	7

Thank you! When you finish please raise your hand and wait for further instructions.

Instructions for the control time (Part 2)

Job Task_Learning session

Please read the instructions carefully and proceed as required.

Today you are employed to do a real job task - a transcription of a recording of a lecture. That means that you are going to listen to a recording of a lecture saved on your computer and write down everything you will hear in the form of a text.

You are allowed to listen to the recording ONLY using the headphones.

There is an open MS Word document **Lecture_Transcript_.doc**. That is the document where you should type the text into.

The recording will be played via VLC Mediaplayer which is now in a background mode on your computer.

Please DO NOT remove or delete ANY files from your computer!

Before you started transcribing let's try out the key combinations, which will help you to manage the recording.

1. Please put the headphones on now.

2. Please press the key combination [*Ctrl+Space*]. Now you should be able to hear the recording.

3. Please press [*Ctrl+←*]. Now you hear that the recording has jumped back a bit. It can be useful if you want to go back and listen to the previous piece of recording once again. To jump forward please press [*Ctrl+→*].

4. Please press [*Ctrl+↓*]. Now you should hear that the recording slows down. It can be easier for you to transcribe the recording when the speed is slower. In order to return it to normal speed please press [*Ctrl+↑*].

5. Finally press the combination [*Ctrl+Space*] again to pause the recording.

   Please DO NOT close the recording or the Word file!

If you have any questions, please raise your hand and we will come up to you to answer.

When you finish please raise your hand and wait for further instructions.

Instruction for the control time (Part 3)

Job Task_Learning session (Cont.)

In the next 10 minutes please do the transcription of the recording. Please transcribe as much recording as possible. **The transcriptions should be word-for-word, which means that everything you hear on the tape should be typed into text.**

These 10 minutes are given to you as a **learning time**. This will allow you to get used to the program and to the usage of key combinations.

Please save your work ONLY by clicking the icon with red exclamation sign **at the upper left corner of your screen**!

We strongly recommend you to save your work at least every 10 minutes.

Once you are ready to start transcribing, please:

---

1. **Click the** **icon** (icon with red exclamation sign at the upper left corner of your screen)**!**

2. Press [*Ctrl+Space*] and start listening to the recording

3. Write down the text you hear into the **Lecture_Transcription_.doc** Use the key combinations when needed.

When 10 minutes are over (**we will announce that**), please:

---

1. Pause the recording by pressing [*Ctrl+Space*].

2. **Save the file by clicking the** **icon** (icon with red exclamation sign at the upper left corner of your screen)**!**

   Please do NOT close the file!

Instructions for the main time (here for the Control group – Group 1)

Job Task_Working Session

Dear participants! Before you proceed with transcribing please read the information related to the purpose of the transcription task you are performing and your beneficiaries.

This study is conducted by the Chair of Organization and Planning in collaboration with the project *Mitschriften Börse* organized at Vienna University. The aim of *Mitschriften Börse* is to provide the students with disabilities better chances of access to higher education. In particular *Mitschriften Börse* assists the disabled students in getting learning materials necessary for completing the courses. There are students (e.g. deaf) who are unable to make lecture notes, which in many cases makes completion of the courses impossible.

To allow such students a better access to the learning materials, *Mitschriften Börse* provides them with the opportunity to get the transcripts of most important lectures. These lectures are first recorded and then have to be transcribed (turned into the typed text) in order to be used by disabled students.

Today you are employed to transcribe a recording of a lecture for the *Mitschriften Börse*.

Your working time is **50 minutes**. Please transcribe as much recording as possible. You should make a **word-for-word transcription** of the lecture you will hear.

Please remember that your work is completely anonymous. Neither your name nor any other information about you will be revealed to Mitschriften Börse.

Once you are ready to start transcribing, please

1. Put on the headphones.

2. **Click the** ![warning icon] **icon!**

3. Press [*Ctrl+Space*] to continue listening to the recording

4. Write down the text you hear into the **Lecture_Transcription_.doc** Use the key combinations when needed.

When 50 minutes are over (**we will announce that**), please:

1. Pause the recording by pressing [*Ctrl+Space*].

2. **Save the file by clicking the** ![warning icon] **icon!**

   Please do NOT close the recording and the Word file!

Instructions for the additional time

Dear participant! Thank you for transcribing a part of the lecture recording!

The obligatory part of your Job Task is finished. However, if you want, you can continue transcribing for **up to 30 minutes**. Please keep in mind **that additional work time is not paid and it won't influence the amount of payment you receive.**

In case you decide to continue, please keep in mind that not only the quantity of transcribed recording, but also the quality of the text which is the output of your effort is of importance for beneficiary. In particular, the text is readable if

- There are as few as possible typos
- There is a clear structure introduced by paragraphs and if necessary numbering

You can use additional time to structure the text you have already transcribed or to continue transcribing the recording.

If you decided to continue transcribing, please

1. **Click the** **icon!**

2. Press [*Ctrl+Space*] to continue listening to the recording

3. Write down the text you hear into the **Lecture_Transcription_.doc** Use the key combinations when needed.

4. When you finish (you have up to 30 minutes), please pause the recording by pressing [*Ctrl+Space*].

5. **Save the file by clicking the** **icon!**

6. **Leave the Word file and recording open! Please do NOT close the recording and the Word file!**

Alternatively you can move directly to the two last questionnaires.

In this case, just leave the Word file and recording open!

Please DO NOT close, remove or delete Lecture_Transcription_.doc or **ANY files from your computer! We will upload the transcription on the webpage of *Mitschriften Börse.***

Questionnaire 3

Please indicate on a 7-point scale (from *not at all* to *very much*), how much you experienced each of these emotions while performing a job:

	Not at all					Very much	
Sympathetic	1	2	3	4	5	6	7
Alarmed	1	2	3	4	5	6	7
Compassionate	1	2	3	4	5	6	7
Disturbed	1	2	3	4	5	6	7
Warm	1	2	3	4	5	6	7
Upset	1	2	3	4	5	6	7
Soft-hearted	1	2	3	4	5	6	7
Troubled	1	2	3	4	5	6	7
Tender	1	2	3	4	5	6	7
Worried	1	2	3	4	5	6	7
Moved	1	2	3	4	5	6	7
Sad	1	2	3	4	5	6	7
Distressed	1	2	3	4	5	6	7
Touched	1	2	3	4	5	6	7

Questionnaire 4

Please indicate on a 7-point scale (from *not at all* to *very much*) whether you agree with the following statements:

	Not at all						Very much
Everybody should help disabled people	1	2	3	4	5	6	7
Usually people help disabled people	1	2	3	4	5	6	7
By providing transcripts I helped the disabled students	1	2	3	4	5	6	7
I find the transcribing task very easy	1	2	3	4	5	6	7
I find the transcribing task very interesting	1	2	3	4	5	6	7
I am satisfied with my performance	1	2	3	4	5	6	7

We are planning to organize an additional session where you can finish transcription of this particular lecture. This session will not be paid. Please indicate how much you are willing to take part in this session.

	Not at all						Very much
I would like to finish the transcription of the lecture recording.	1	2	3	4	5	6	7

Please, provide here the following information about yourself

Age_____   Gender (please tick):   ☐ male / ☐ female

Semester at the University_____ Years of work experience_____

Language knowledge (please tick)

German   ☐ native / ☐ advanced / ☐ intermediate / ☐ beginner

English   ☐ native / ☐ advanced / ☐ intermediate / ☐ beginner

Did you know about *Mitschriften Börse before this study*   ☐ yes / ☐ no

Have you ever taken part in *Mitschriften Börse before this study*   ☐ yes / ☐ no

Are you going to take part in *Mitschriften Börse after this study*   ☐ yes / ☐ no

As stated earlier, your responses to all of the questionnaires are confidential.

In return we would ask you to maintain confidentiality about the procedures used in this study since any pre-knowledge can influence the participant and the data collected cannot be used for the research.

If you have any complaints, concerns, or questions about this research, please feel free to contact us via email **experiment.orga@univie.ac.at.**

Thank you very much for your participation!

List of additional variables introduced within the main experimental procedure

Name of the variable	Description of the variable	Measurement of the variable
SVO_Grad	Social orientation in °	Measured according to the SVO slider. Measurement conducted in the very beginning of the experiment.
SVO_Lable	Social orientation coded as "0" for "Individualistic" and as "1" for "Prosocial"	
Empathic trait	Empathy measured as Trait	Measured according to the Empathy quotient. Measurement conducted in the very beginning of the experiment.
Empathic state	Empathy measured as State	Measured according to the Empathy measure. Measurement conducted at the very end of the experiment.
Distress state	Distress measured as State	
Prescriptive norm	Level of agreement with the statement "Everybody should help disabled people"	Measured using 7-point Likert Scale. Designed as a proxy for awareness of the prescriptive norm.
Descriptive norm	Level of agreement with the statement "Usually people help disabled people"	Measured using 7-point Likert Scale. Designed as a proxy for awareness of the descriptive norm.
I helped	Level of agreement with the statement "I helped the disabled students"	Measured using 7-point Likert Scale.
Task easiness	Level of agreement with the statement "Transcribing task was easy"	Measured using 7-point Likert Scale.
Task interesting	Level of agreement with the statement "Transcribing task was interesting"	Measured using 7-point Likert Scale.
Satisfied	Level of agreement with the statement "I am satisfied with my performance"	Measured using 7-point Likert Scale.
Want_finish	Level of agreement with the statement "Want to finish"	Measured using 7-point Likert Scale.
Age	Age of the participant in years	
Gender	Gender of the participant.	Coded as "0" for females and "1" for males
Semester	Semester at the University	
Experience	Work experience in years	
German	German language proficiency.	Coded as "1" for "Beginners", "2" for "Intermediate", "3" for "Advanced" and "4" for "Native"
English	English language proficiency.	Coded as "1" for "Beginners", "2" for "Intermediate", "3" for "Advanced" and "4" for "Native"
Know_MB	Answer to the question "Did you know about MB before".	Coded as "0" for "No" and "1" for "Yes"
Takepart_MB	Answer to the question "Have you taken part in MB".	Coded as "0" for "No" and "1" for "Yes"
Will_takepart_MB	Answer to the question "Are you going to take part in the MB".	Coded as "0" for "No" and "1" for "Yes"

Tests of normality for the main output variables in the main time

Parameter	Group	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Preciseness	1	,205	27	,005	,856	27	,001
	2	,277	25	,000	,716	25	,000
	3	,283	24	,000	,720	24	,000
	4	,254	24	,000	,802	24	,000
Mistakes	1	,192	27	,012	,871	27	,003
	2	,322	25	,000	,740	25	,000
	3	,194	24	,020	,798	24	,000
	4	,243	24	,001	,700	24	,000
Speed_Min	1	,087	27	,200[*]	,985	27	,954
	2	,277	25	,000	,625	25	,000
	3	,325	24	,000	,757	24	,000
	4	,264	24	,000	,643	24	,000
Speed_Words	1	,104	27	,200[*]	,982	27	,912
	2	,155	25	,123	,959	25	,397
	3	,142	24	,200[*]	,928	24	,090
	4	,113	24	,200[*]	,954	24	,327

Instructions for the follow-up study (Part 1)

Dear Student,

You are now taking part in the second stage of the study in Organisational behaviour. The first stage was organised in May 2013. In that study your main task was to transcribe a recording of the lecture. You had a limited amount of time and were paid a fixed remuneration for your work.

This time we want to deepen our understanding of those decisions that you took in the original study. To achieve this goal we provide you with two assignments. First one relates to your decision-making process during the original study and second one constitutes in evaluating the output of other participants (both the transcripts provided and your evaluation of them is completely anonymous).

You have around 15 minutes for completing each assignment, which adds up to 30 minutes of participation time. However, time limits are not fixed and if necessary, you can stay longer. In any case your fixed remuneration for completing both assignments is 10 Euro.

We kindly ask you to read the instruction carefully! Please feel free to ask any questions relevant for performing your tasks!

Assignment 1

In the original study (May 2013) you were transcribing the recording of the lecture. There were several parameters, which could be more or less important for your final output, for example:

- amount of the recording transcribed (how much of the lecture one transcribed),
- closeness of the transcript to the recorded lecture (word-for-word transcript or not),
- accuracy of the typing (misspellings, using of capital letters, punctuation, etc.)
- formatting of the text (paragraphing, etc.)
- other parameters.

Which parameters, do you think were more important and which less important for doing a good job?

Please fill in the table. Thinking about relative importance, you can imagine for example, that if all parameters have equal importance, then the figures in the column can look like 25 / 25 / 25 / 25 or if some parameters are more important than others, then the figures in the column can look like 10 / 20 / 30 / 40 or 40 / 20 /20 /20 or any other combination which sums up to 100%.

Relative importance of different parameters of output

Parameter	Relative importance of parameters (in %)
Amount of the recording transcribed (transcribing as much as possible)	
Closeness of the transcript to the recorded lecture (making word-for-word transcript)	
Accuracy of typing (providing text free from misspellings, applying capitalization to the nouns, punctuation)	
Formatting (applying paragraphing, text alignment, spacing, etc.)	
	100%

Where there any other parameters important for you? (Please write in the space provided)

_____

_____

All participants of the original study received the information that they perform the task for the Mitschriften Börse project of the Vienna University, which aim is to help the disabled students to get access to lecture materials necessary for their study.

Besides the information about the Mitschriften Börse, some randomly chosen participants (may be you were one of them) got the additional information about a particular customer – a person who has applied to Mitschriften Börse to receive the transcripts of these lectures. These participants received the following additional information:

Today you are employed to transcribe a recording of a lecture for Alexander Friedrich, who is one of the students registered within the Mitschriften Börse.

Alexander is 28. He is a student of the Faculty of Sociology of Vienna University. This semester he has applied to the Mitschriften Börse asking for the transcripts of the lectures in Sociology.

Due to the injury that he has suffered several years ago, he is unable to make the lecture notes himself.

Please think about potential difference in effort for you and other participants **in case if provided with additional information about particular customer**, as above (Please tick the relevant box. Only one answer in each column is possible):

My effort would be		**Effort of other participants would be**	
Higher	☐	Higher	☐
Same	☐	Same	☐
Lower	☐	Lower	☐

Some randomly chosen participants were asked to provide their names to the customer (may be you were one of them).

Please think about potential difference in effort for you and other participants in **case if asked to provide the name to the customer** (Please tick the relevant box. Only one answer in each column is possible):

My effort would be		**Effort of other participants would be**	
Higher	☐	Higher	☐
Same	☐	Same	☐
Lower	☐	Lower	☐

Thus, participants had some difference in information provided about the customer and to the customer. As a result the participants assigned different importance to different parameters of output (for example, some could believe that it is important to transcribe as much as possible, while quality of the transcript does not matter; others could try to make the transcript as close as possible to the original lecture, even if it meant to transcribe less of the recording).

Please think about relative importance of different parameters for participants with different information. What parameters could have been more important for participants with information about particular customer? Were participants paying more attention to some specific parameters, in case they had to provide their name to the customer?

Fill in the table below, remembering that the sum of percentages in each column should equal 100%.

We stress that for filling in this table it is essential to put yourself in the shoes of other participants. Think about what parameters of output THEY (on average) could consider important, taking into account the information they had! **THE PARTICIPANT CLOSEST IN HIS/HER EVALUATION TO THE REAL OUTCOME OF OUR ORIGINAL STUDY WILL WIN A 30 EURO AMAZON VOUCHER!**

Relative importance of different parameters of output for different groups of participants

Parameter	Relative importance (in %) of output parameters for the group of participants with following information*			
	No info about particular customer Name is not provided to the customer	**Have info about particular customer** **Name is not provided to the customer**	No info about particular customer **Asked to provide name to the customer**	**Have info about particular customer** **Asked to provide name to the customer**
Amount of the recording transcribed (transcribing as much as possible)				
Closeness of the transcript to the recorded lecture (making word-for-word transcript)				
Accuracy of typing (providing text free from misspellings, applying capitalization to the nouns, punctuation)				
Formatting (applying paragraphing, text alignment, spacing, etc.)				
	100%	**100%**	**100%**	**100%**

*All participants knew that they worked for Mitschriften Börse.

This is the end of the Assignment 1. Thank you. Please proceed to the Assignment 2.

Instructions for the follow-up study (Part 2)

Assignment 2

In this assignment you need to read four transcripts of the same lecture and share with us your opinion about each of them.

Transcript 1

Please read the Transcript 1 and indicate on a 7-point scale (from *not at all* to *very much*), how true each of these statements:

Transcript 1 is easy to read

Not at all   Very much
    1   2   3   4   5   6   7

Text in Transcript 1 looks like a word-for-word transcript of the lecture

Not at all   Very much
    1   2   3   4   5   6   7

Transcript 1 has high quality of typing (free from misspellings, capitalization in the nouns, punctuation)

Not at all   Very much
    1   2   3   4   5   6   7

Transcript 1 has high quality of formatting (paragraphing, text alignment, spacing, etc.)

Not at all   Very much
    1   2   3   4   5   6   7

The overall quality of this Transcript 1 is high

Not at all   Very much
    1   2   3   4   5   6   7

Transcript 2

Please read the Transcript 2 and indicate on a 7-point scale (from *not at all* to *very much*), how true each of these statements:

Transcript 2 is easy to read

Not at all   Very much
    1   2   3   4   5   6   7

Text in Transcript 2 looks like a word-for-word transcript of the lecture

Not at all   Very much
    1   2   3   4   5   6   7

Transcript 2 has high quality of typing (free from misspellings, capitalization in the nouns, punctuation)

Not at all   Very much
    1   2   3   4   5   6   7

Transcript 2 has high quality of formatting (paragraphing, text alignment, spacing, etc.)

Not at all   Very much
    1   2   3   4   5   6   7

The overall quality of this Transcript 2 is high

Not at all   Very much
    1   2   3   4   5   6   7

Transcript 3

Please read the Transcript 3 and indicate on a 7-point scale (from *not at all* to *very much*), how true each of these statements:

Transcript 3 is easy to read

Not at all   Very much
    1   2   3   4   5   6   7

Text in Transcript 3 looks like a word-for-word transcript of the lecture

Not at all   Very much
    1   2   3   4   5   6   7

Transcript 3 has high quality of typing (free from misspellings, capitalization in the nouns, punctuation)

Not at all   Very much
    1   2   3   4   5   6   7

Transcript 3 has high quality of formatting (paragraphing, text alignment, spacing, etc.)

Not at all   Very much
    1   2   3   4   5   6   7

The overall quality of this Transcript 3 is high

Not at all   Very much
    1   2   3   4   5   6   7

Transcript 4

Please read the Transcript 4 and indicate on a 7-point scale (from *not at all* to *very much*), how true each of these statements:

Transcript 4 is easy to read

Not at all   Very much
    1   2   3   4   5   6   7

Text in Transcript 4 looks like a word-for-word transcript of the lecture

Not at all   Very much
    1   2   3   4   5   6   7

Transcript 4 has high quality of typing (free from misspellings, capitalization in the nouns, punctuation)

Not at all   Very much
    1   2   3   4   5   6   7

Transcript 4 has high quality of formatting (paragraphing, text alignment, spacing, etc.)

Not at all   Very much
    1   2   3   4   5   6   7

The overall quality of this Transcript 4 is high

Not at all   Very much
    1   2   3   4   5   6   7

Imagine, that you are a recipient of the transcript. Which transcript would you prefer to have (only one answer is possible)?

Transcript 1 □
Transcript 2 □
Transcript 3 □
Transcript 4 □

Now when you have read all four transcripts, please think which parameters could be more important for the recipient of the transcript and which less important. Please fill in the table below.

Relative importance of different parameters of output

Parameter	Relative importance of parameters (in %)
Amount of the recording transcribed	
Closeness of the transcript to the recorded lecture (word-for-word transcript)	
Accuracy of typing (text free from misspellings, capitalization in the nouns, punctuation)	
Formatting (paragraphing, text alignment, spacing, etc.)	
	100%

Are there any other parameters which can be important for the customer? (Please write in the space provided)

_____

_____

This is the end of the Assignment 2 and the end of the session.

Thank you very much for sharing your opinion with us! All the information you provided us with is seen as strictly confidential and will be used only for the research purposes.

In return we would ask you to maintain confidentiality about the procedures used in this study since any pre-knowledge can influence other participants and the data collected can't be used for the research.

We also want to thank you on behalf of Mitschriften Börse project and express our hopes that you will again participate in similar projects both within and outside of the University.

If you have any complaints, concerns, or questions about this research, please feel free to contact us via email **experiment.orga@univie.ac.at.**

We would also like to offer you the opportunity **to earn additional 10 Euro for evaluating another two sets of transcripts (4 transcripts in each set).** Please let us know if you are willing to participate and when. We can either provide you with transcripts for evaluation right now or you can register for performing this job on 27 March, 2014 between 16:00 and 19:00.

Abstract (English)

Complexity of tasks, environmental uncertainty and incomplete contracting are only a few prominent characteristics of modern jobs. While these factors make the traditional (pecuniary) incentives more difficult to apply, at the same time they open up the room for prosocial motives to govern behavior.

In our research we focus on the ways of motivating prosocial behavior of employees towards customers. Based on the results of interdisciplinary literature review, we focus on the identification in the employee-customer dyad as one of the promising mechanisms working in this direction. We distinguish between two ways of identification: making customer identifiable by employee (by providing employee with the information about a single customer in form of a photo, a personal story) and employee to the customer (by opening the name of the employee to the customer).

We hypothesize that by making customer identifiable one can decrease social distance between the employee and the customer and, through a mechanism of empathic concern and perspective taking, increase both the general level of effort and customer-oriented focus in distribution of this effort across attributes of the task. By making the employee identifiable one can increase observability of employee's output, which, via a mechanism of image concerns, can lead to the increase in the general level of effort, but with self-image focus in distribution of this effort across attributes of the task.

To test these hypotheses we conducted a laboratory experiment, where subjects worked under condition of incomplete contracting and performed multiattribute task for a real customer. 100 students of the University of Vienna took part in the experimental sessions; and more than 30 of these students also participated in the follow-up study, which provided us with a measure of customer satisfaction and contributed to a more profound understanding of the behavior observed in the experimental study.

As hypothesized, exposure to the information about individual customer led to the increase in overall level of effort with trade-off between parameters of output moved towards satisfying customer's needs. Opening employee's identity to the customer, as expected, crowded in the image-related motives and led to the trade-offs strongly shifted towards parameters of output signalling effort of the employee, but not meeting the needs of the customers. As a result the output produced by employees exposed to the customer enjoyed higher satisfaction rates among potential customers.

Thus, readiness of the employees to exhibit prosocial behavior towards their customers can be increased by using such specific contextual tool as identification, but its application should be careful and take into the account both the individual differences among employees and specifics of the task itself. Future research is necessary to provide clear guidelines in this direction.

Abstract (German)

Komplexität von Aufgaben, umgebende Unsicherheit und unvollständige Verträge sind nur wenige prominente Faktoren, die es erschweren traditionelle Motivationsmechanismen in Unternehmen anzuwenden. Währenddessen erlauben uns die gleichen Faktoren die Vorzüge von pro-soziale Motiven zu erkennen, um Verhalten entsprechend zu steuern.

Basierend auf interdisziplinärer Literaturrecherche unterscheiden wir zwischen pro-sozialen Verhalten, das einerseits durch andere und andererseits durch selbstbezogene Motive ausgelöst werden kann. Das durch andere ausgelöste pro-soziale Verhalten wird vielfach mit Empathie in Verbindung gebracht und das selbstbezogene pro-soziale Verhalten wird mit der Wahrung des Selbstbildes assoziiert. Bestehende Forschung zeigt, dass Empathie zunimmt, wenn die Distanz in einem pro-sozialen Akt zwischen Bereitsteller und Begünstigen geringer wird. Anderseits steigt auch die Wahrung des Selbstbildes, wenn pro-soziales Verhalten beobachtbar wird.

In unserer Forschung fokussieren wir auf das pro-soziale Verhalten von Angestellten gegenüber Kunden. Wir stellen die Hypothesen auf, dass es zu einem Anstieg der Leistung kommt, wenn die soziale Distanz zwischen den Angestellten und den Kunden reduziert wird bzw. wenn die Angestellten identifizierbar gegenüber dem Kunden sind. In mehrdimensionalen Aufgaben, abhängig von verwendeten Mechanismen, kann die Leistung hinsichtlich des Outputs in Form von Quantitäts- oder Qualitätsattributen gelenkt werden.

Damit diese Hypothesen getestet werden können, wurden 100 Studierende der Universität Wien angeworben, eine Aufgabe für einen echten Kunden durchzuführen.

Wie angenommen, sorgten die Teilnehmenden, deren Identität offen gegenüber dem Kunden war, für eine positive Imagedarstellung und maximierten die sichtbaren Attribute der Qualität. Obwohl die Quantität für den Kunden sehr wertvoll ist, aber auch schwer messbar, wurde die Leistung somit auf Kosten der Quantität sichtbar gemacht.

Teilnehmende, die individuelle Informationen über den Kunden erhielten, maximierten die Quantität in Verbindung mit mehreren Qualitätsattributen. Gemäß den Ergebnissen der weiterführenden Studie, dank dieser kundenzentrierten Strategie, erhielt deren Output signifikant höhere Zufriedenheitsraten bei den potentiellen Kunden.

Curriculum Vitae
Oksana Galak

Education

2009 – University of Vienna, PhD in Management

2001, Tashkent Financial Institute, Credit Faculty, **Master's** degree (honours)

1999, Tashkent Financial Institute, Credit Faculty, Banking Department, **Bachelor's** degree (honours)

Work History

2003 – 2009 Tashkent Financial Institute (TFI), Uzbekistan, Banking Department, Lecturer (from 2007 – Senior Lecturer).

Taught courses:

- Banking theory
- Management and Marketing in Banking
- Monetary Policy
- Risk management in Banking
- Banks, Money and Credit, etc.

2007 – 2008 Office for entrepreneurial activities support, TFI, Head of the Office

- Development of the Office's strategy, administration of the Office's daily activities, personnel management
- Cooperation with foreign academic institutions, including University of Saarland (Germany), University of Alicante (Spain), St-Petersburg Technological University (Russian Federation).
- Dissemination of the results of Office's activity

2005 – 2007 Joint European Tempus Project JEP 25085-2004, SEASCAPE, Member of the Office Staff

- Organisation of training, seminars and various event (e.g. business plan competition)
- Financial management
- Coordination of liaison and correspondence between parties of the project, including between TFI and Project coordinator University of Saarland, Germany

2003 – 2004 JSC Transformator, International relations and Investment Department, Senior associate

- Marketing, analytical research,
- Working on the project on Restructuring of the Plant: workflow, business planning
- Keeping correspondence and arranging meetings with foreign clients, translation

---

2002 – 2003 Ministry of Macroeconomics and Statistics of Uzbekistan, Research Centre, Senior associate

- Marketing, financial and economic appraisal of the investment projects
- Conducting Market analytical reviews of the local and international markets
- Feasibility study of textile industry projects requiring cotton provision from government stock

2000 – 2002 National Bank of Uzbekistan, Project Finance Centre, Marketing department, Senior associate

- Marketing reports within the framework of feasibility appraisal of the investment projects financed by NBU
- Maintaining financial and statistical databases and providing the management of the Bank and Government authorities with data concerning investment activities by NBU

1997 – 1999 Tashkent Financial Institute, International Relations department, Interpreter / Assistant

Special projects

- 2008 - Business Fair at Tashkent Financial Institute – Organising event on executive level
- 2003/2004 - Project of purchasing of the equipment within the framework of Uzbek - Japanese government grant – Legal issues, wide scope of translation and administration activities
- 2001 – Teaching at National Bank's "Centre for qualification upgrading" – "Marketing as a part of monitoring process of the banks"
- 1997 / 1999 - Assistant for the Fulbright Fellowship members teaching at Tashkent Financial Institute. Wide range of translation and organisational activities.

Main Publications

- "Higher education and development of small business and entrepreneurship", Conference paper, 2008
- "Retail deposits in Uzbekistan: contemporary market analysis", Eastern Business bulletin, 2008
- "Retail strategy of the commercial bank ", Conference paper, 2007
- Training manual on "Money, credit and banks" (with Prof. Dr. S.Arzumanyan), TFI, 2007.
- "Establishment of the new type of the financial system: policy recommendations for Central Asian countries", Conference paper, 2006
- Training manual on "Management and marketing in Banking" (with Prof. Dr. Sh.Abdullaeva, et al), TFI, 2006.
- Training manual on "Risk management in Banking" (with Prof. Dr. S.Arzumanyan), TFI, 2006.

Languages

Russian (native), English (fluent, IELTS 7.5), German (very good, certified level B2), Uzbek (basic)

Contact details

e-mail: oksana_galak@mail.ru

https://univie.academia.edu/OksanaGalak