Vladyslav Romashov
34760

How response format artificially increases emotion recognition rates

Master Dissertation prepared
under supervision of
Dr. Krzysztof Przybyszewski

Warsaw 2018

Author of the dissertation: Vladyslav Romashov

Title of the dissertation: How response format artificially increases emotion recognition rates

Abstract

In this dissertation, I have outlined theoretical and empirical background of emotion recognition studies, their application in academic and applied context. Different methodological issues of emotion judgement studies discussed and one – response format studied in the research section. This Dissertation unfolds as follows: first part contains short history of study of facial expression, measument methods of facial behavior and description how knowledge about facial expression can be applied in business context; second part is devoted to methodological issues in subjective measurement approach. Third part is empirical part, where method, results and conclusions of Dissertation's author experiment described. Hypotheses were confirmed. Results once again show that Forced Choice format indeed artificially increases agreement. Future researches should use Fixed Choice, which is as easy to apply as Forced Choice, but which has lesser forcing effect.

Key words: unbiased hit rate, fixed choice, forced choice, judgment study, facial expressions, agreement

Field code in the program "Erasmus for all": 14400

Title of the dissertation in Polish: Jak format odpowiedzi sztucznie zwiększa dokładność rozpoznawania emocji

Table of contents

1. Introduction

Nowadays because of globalization our possibilities are raising. New technologies give us new ways to travel faster, communicate with anyone around the globe. This process of connecting people from different countries, ethnicities and cultures raises our requirements for understanding others not only in personal communication but in business domain as well. Indeed, traveling and having vacations, working in many countries around the globe demands better understanding of others. If we are interested to communicate our intentions properly to those who are visitor or colleagues from abroad as well as from our country, we must understand them better. One way how we can achieve it is to get deeper insight in their culture and traditions. Another way and easier one is to understand their nonverbal behavior. Specifically, facial expression of emotions, which is believed to be universal. In this section, I will beefily give answers for the following questions: How universals and culture specifics in facial expression were studied? How we can measure facial behavior? How we can use facial behavior in business context, especially in marketing researches as well as in deception detection? I will start with "Historical Sketch" which will give you some insight in most prominent studies to date and basic concepts of the field.

1.1. Historical Sketch

The review of research history on the universality of facial expression of emotions in most articles begin with Charles Darwin's book «The Expression of the Emotions in Man and Animals» (1872), in which he proposed universality thesis, the idea that at least some emotion have universal expression in all human beings (but see Russell, 1994). However, because of predominant belief in end of XIX and beginning of XX century, called "relativism", that expression of emotion is different in cultures all around a globe, and treated as a language, that is to understand someone's facial expression of emotion from another culture, you should learn it first, his proposal was unpopular and/or rejected and forgotten soon. After almost one century, the new wave of researches in the field was encouraged by publication of two Silvan Tomkins' books (1962, 1963), where he after Darwin expanded idea of universality. Then, Paul Ekman (1969) and Carrol Izard (1971) found in two separate studies each, that participants had strong cross-cultural agreement in labelling six or ten (Ekman's and Izard's study, respectively) expressions in number of countries, both Western and Non-Western. With one exception, that could be bullet point in question of universality thesis, but Paul Ekman (Ekman, 1985) found gap in these studies, which relativists could use to question these results. All participants in theirs

studies were from literate, industrialized countries and because studies took place in sixties, era of TV, they could have learned to understand western-expressions through watching Western media. To fill that gap, Paul Ekman and colleagues (1971) conducted study in preliterate settlements in Papua New Guinea, who had almost no contact with rest of the world and thus couldn't have possibility to learn to understand these expressions. Results were highly similar. Both, adults and children, had strong agreement in labeling all emotion significantly higher than chance. Similarly, expressions of those preliterate people were highly recognized by American students (Ekman, 1973). Since that, hundred studies have replicated those results and expand our knowledge about universality of emotion. In 1986 Ekman added "contempt" emotion into list of universal. Up to date, there are evidence that other emotions or affects, beyond "standard six", have universal facial expression (Cordaro, 2014; Keltner, Tracy, Sauter, Cordaro, & McNeil, 2016; Matsumoto, Keltner, Shiota, O'Sullivan, & Frank, (2008), but there is lack of agreement between researcher which emotions have universal expression (Ekman, 2016). However, evidence of universality comes not only from high cross-cultural labeling agreement, but from studies of measuring facial behavior when emotions elicited, in both sighted and blind people, and nonhuman primates; correlation between emotional facial expression and appraisals, physiology, and subsequent behavior (Matsumoto et al., 2008). Some evidence also suggests there is genetic factors on facial expression, but literature on that question is too scarce. Papadatos and colleagues' (Papadatos, Alexiou, Nicolopoulos, Mikropoulos, & Hadzigeorgiou, 1974) study showed that hypoplasia of lip corners depressor muscles is familial, monozygotic demonstrate have more similarities than dizygotic twins in development of social smiles, both timing and timing of fear reaction during first year of life, have more similarities in eye-blink startle reactions (Carlson, Katsanis, Iacono, & McGue, 1997), congenital blind individuals showed more similarities with their sighted relatives for some facial movements (Peleg, Katzir, Peleg, Kamara, Brodsky, Hel-Or, Keren, & Nevo, 2006) and finally, most recent study suggests twins resemblance within emotional facial expression, and varied with emotional category and were more pronounced for positive than negative emotions (Kendler, Halberstadt, Butera, Myers, Bouchard, & Ekman, 2008)

Beyond facial expression, there are a bunch of studies showing evidence for universality of expression of emotions in other communicative channels (see Keltner et al., 2016 for review), including posture and gestures (f.e. Wallbott, 1998), voice (vocal expression; Scherer, 2003, Juslin & Laukka, 2003), vocalization (Cordaro, 2014; Cordaro, Keltner, Tshering, Wangchuk & Flynn, 2016; Sauter, Eisner, Ekman & Scott,

2010), physiology (f. e., Levenson, 2003), touch (Hertenstein, Keltner, App, Bulleit, & Jaskolka, 2006; Hertenstein, Holmes, McCullough, & Keltner, 2009) and music (f.e. Juslin & Laukka, 2003).

However, universality of expression is one side of the coin. For example, Paul Ekman's theory of emotion (Ekman, 1984, 1992a, 1992b, 1994a, 1994b, 1997, 1999; Ekman & Cordaro, 2011) sees that "emotion is not a single affective or psychological state, but family of related states" (pp364, Ekman & Cordaro, 2011). Term "theme" used to describe comprised of the characteristics unique to family, which shares all members of that family. Themes not only common feature of emotion, but the core element, product of evolution, essential in all emotions, its phylogenetical part. Variations refer to differences of members around the family. Variations are responsible for individual differences, the result of social experience or learning (Ekman, 1992a, 1997; Ekman & Cordaro, 2011). He described his view with anger family, where intensity variations could be between *annoyance* and *rage*. Other examples could be following: *resentment*; *indignation* and *outrage*; *vengeance*; *berserk*. Those in italics are members of one family – anger one, but they are different forms of anger (Ekman, 1997). This view is close to Shaver et al (1987) prototypical view on emotions, but they do not use term families.

As can be seen, social experience or learning can influence on emotions, as well as culture could. "Culture, though, are not geo-political states; they are socio-psychological entities" (pp130, Matsumoto, 1991). It has been found, that people have "display rules" – rules learned early in life which dictates and helps people to control and/or modify their own facial expression depending on social circumstances (Matsumoto, 1998; Matsumoto et al, 2005). When expression of emotion deliberately controlled, so person shows modified expression in place of expression of felt emotion, we are dealing with influence of display rules (Ekman, Friesen & Ellsworth, 1972). Display rules defines appropriateness of expressive behavior; assumed that people can control their expressive behavior according with social rules (Ekman & Friesen, 1975).

Display rule vary with culture, for example, in Matsumoto et al. (2005) study, participants from Russia and USA thought it is more appropriate to show expressions of anger and contempt in comparison to Jappanies participants, and expression of joy was most appropriate emotion in every studied country. Another study showed that this difference exists even in borders of same geo-political country with different ethnical groups (Matsumoto, 1993). Moreover, there is correlation between expressivity and individualism, thus more individualistic cultures are more expressive (Matsumoto, Yoo, & Fontaine, 2008)

One thing what we "believe" we should show, and other thing is how culture influence on our perception of emotion in others. There are burden of studies shows that exist some difference in emotion recognition accuracy between cultures. One of the earliest studies, done by Matsumoto (1992) showed that participants from Japan had significantly lower accuracy for anger, disgust fear and sadness, but the same accuracy for remaining emotions – happiness and surprise. More interesting result of this study is that this effect was not only for same rater-actor culture, in other words – American participants had greater recognition for those four emotions than Japanese, regardless of actor gender or culture. However, there are some studies proving in-group advantage in emotion recognition accuracy (Elfenbein, 2015; Elfenbein, & Ambady, 2002a, 2002b, 2003a, 2003b; but see Matsumoto, 2002, 2007)

People in different cultures not only recognize emotions differently, but also perceive of expression in different way. In Matsumoto and Ekman's study (Matsumoto & Ekman, 1989) participants firstly rated intensity of each emotion and then intensity of expression itself, which eliminated possibility of differences between two terms for expression due to translation. Americans rated all but one (disgust) emotions as expressed more intense than Japanese did, regardless of first or second rating. Although, participants from different cultures rate expressions intensity differently, it does not have influence on their accuracy (Matsumoto, 1992).

Another difference between ethnicities could be occurred in frequency how often people experience different emotions. There are evidences, that African Americans experience anger more often than European, Asian or Latino Americans in USA (Matsumoto, 1993)

Finally, language can influence on emotion recognition accuracy, even within same person. In Matsumoto & Assar's (1992) study, participants from India judged 40 photos of five emotions both by selecting single term as well as intensity rating for each emotion. Participants firstly had done this either in Hindi or English, and after two weeks, during session in the remaining language. Three out of five emotions were statistically more readily recognized in English than in Hindi (anger, fear and sadness), and there was gender-language interaction. For example, male judges recognized emotions better in English language, but not females.

1.2. How can we measure emotions in the face?

Before proceeding to section how we can use facial expression of emotion, we have to understand how we can measure emotions in the face. Up to date, there are three most popular methods to understand facial expression: electromyography (EMG), manual or

automated coding and observer-based approach (Cohn & Ekman, 2005; Matsumoto & Hwang, 2016). For further simplification, we will separate those three methods into two broad categories – objective and subjective approaches.

1.2.1. Objective measurement

1.2.1.1. Electromyography (EMG)

During EMG measurement, researches attach electrodes to participants face and record muscle contraction based on detection of minute electrical discharges from the contracted muscle. The main advantage of EMG measurements is that can detect very low contraction of the muscle, which even do not create observable appearance changes. However, EMG measurements can't record contraction of single muscle, but measure "regions" (Fridlund, 1991). Before measurement, face should be prepared (with paste) for improvement conductance which can focus participants' attention to the face. Additionally, because of electrodes facial behavior could be restrained to prevent falling of electrodes. Finally, because of placement of electrodes, facial EMG often measure facial behavior only on one side of face, which gives no opportunity to measure symmetry of the expression, the sign of expression sincerity (Ekman & Rosenberg, 2005)

Moreover, EMG studies require expensive equipment and because of that, EMG measurement is a very rarely used in business context.

1.2.1.2. Manual and automatic coding

Between 1924 and 2002 years were developed at least 14 techniques to measure facial behavior (Cohn & Ekman, 2005). Most often used techniques used todays are Facial Action Coding System (FACS; Ekman, Friesen & Hager, 2002a, 2002b) and Maximally Discriminative Facial Movement Coding System (MAX; Izard, 1979; Matsumoto & Hwang, 2016). The first one is comprehensive anatomically-based system for measuring all possible facial behavior and some addition behavior like head positions, gaze direction and gross movements like shrugs or nods. MAX instead is theoretically based technique and could be used only to measure emotions it was intended to measure.

Because FACS is anatomically-based system it could be more applied in variety situations. For studying emotion expressions, codes obtained previously could be studied for correspondence with codes theoretically or empirically related to emotions and if correspondence is high, researcher can be sure he or she is dealing with emotion expression. But FACS could be used for studying not only the emotion expression, but any facial behavior. This comprehensiveness makes FACS highly time consuming. For coding one minute of facial behavior it takes up to one hour of time (Ekman & Rosenberg,

2005). To remedy this drawback, for at least two decades there are ongoing developing of automated facial coding (AFC).

Nowadays several companies develop AFC based on FACS and give promising results. For example, FaceReader software developed by Noldus Information Technologies has almost reached minimal required score for passing FACS certification exam – 0.70. In fact, for WSEFEP (Olszanowski, Pochwatko, Kuklinski, Scibor-Rylski, Lewinski, & Ohme, 2015) and ADFES (Van der Schalk, Hawk, Fischer, & Doosje, 2011) stimuli sets, agreement between FaceReader and manual coding were 0.70 and 0.68, respectively (Lewinski, den Uyl, & Butler, 2014). Thus at least for one set it has already passed exam. But scientific society still consider automated coding not enough developed (Matsumoto & Hwang, 2016; for recent review of AFC interested readers are referred to Buck & Miller, 2016).

Behavior coding technique is highly time-consuming when used manually and automatic behavior coding is still not enough developed, thus making application rare in field. That is why measuring facial behavior using coding systems is not easy task. It requires trained coders, and highly time-consuming. Are there any other ways to study emotions in marketing beyond direct measuring of thee face?

1.2.2. Subjective measurement

In subjective measurement, panels of observers watching facial behavior and decide what it does express. Thus, inference about internal state is based on other's perception. This type of studies is also known as judgment studies.

Both methods, objective and subjective could be used to answer the same question, but still there are the difference too. Let's consider the following example. Some company develops ad for their new refreshing drink. In their ad, little boy sips a bit of this drink and then smiles. To answer some questions, we can use both methods. However, if we are interested in understanding how people perceive if kid is pleased after tasting the drink, we would not use behavior coding approach, observer-based approach required. Although, objective-behavior coding approach could show that boy is smiling, which is sign of positive feeling, people may see him not so pleased, for example, because of context (Barrett, Lindquist & Gendron, 2007)

Those two methods differ how we treat facial behavior. In case of behavior measuring approach, facial expression is treated as reaction, while in observer-based approach it treated as a stimulus (Ekman, Friesen & Ellsworth, 2013).

We can use both approaches to study facial behavior and if discrepancy emerges, there are some possible explanations (Cohn & Ekman, 2005):

- Behavior coding approach show positive, while observer-based negative. In that case people can just don't know where to look at, or facial behavior is too subtle or fast to be seen by naked eye.

  Researches often implies erroneous conclusion that facial expression does not have it, when people can't recognize it in the face, however, those results could be explained by other factors, such as mentioned above (Ekman et al., 2013).
- Observer-based approach show positive, while behavior coding negative. Behavior coding technique used in the study is not comprehensive enough (for example it does not measure blushing, useful sign of embarrassment). Other explanation possible too, for example facial measurement was not reliable or standardized.
- Negative results with behavior coding and observer-based approaches. In those cases, phenomena studied just could not be expressed in the face.

There is fine example how two approaches could lead to different results. In 2013 Hehman and colleagues (Hehman, Leitner & Gaertner, 2013) demonstrated that the same facial expression – neutral (none of muscles is contracted) could be perceived as more or less aggressive depending on head tilt forward or backward. If the head tilt is not considered, then behavior measurement approach would show that in all instances facial behavior was the same – neutral, while observer-based approach shows that very same facial behavior could be perceived differently. Those who planning studies must consider that in preparation and measurement stages of future research.

1.3. How can we use facial expressions in business?

Expressions of emotion – is a part of nonverbal behavior, which signals internal feelings and motives of person. Ability to recognize expressions of emotions in people around us is useful skill, which helps to build bridges with other members of family, friends and business partners, understanding their feelings and interact with them. This ability to recognize emotions can be useful in any life domain, starting from close relationship to business. In the last one, scrutiny analysis of behavior might give us information not only about emotions itself (for example, anger), other affects (for example, pain), or reflexes (for example, startle), but also about illustrators (for example, underling most important point in sentence), or facial emblems (for example, disbelief), display cognitive load (frown brows or slowed blinking) or signaling intentions (for example, facial expression of imminent aggression; Matsumoto & Hwang, 2014; Romashov, 2016), and even can betray a lie (micro or subtle expressions; Ekman, 1985).

Any business employee can benefit with high skill in emotion recognition, both average executive and superiors, especially in service providing sphere. Salesman can track customer's reactions to own's offers, bank clerks can conduct preliminary screening of unreliable borrowers, HR personnel can get deeper insight interpersonal conflicts, superiors can better understand employees and adequately predict their performance.

Beyond that, there are two spheres where expression of emotion can give the most fruitful results. Facial expression can be a useful source of information in marketing research and when we are trying to catch a liar.

1.3.1. Emotions in marketing research

Measuring emotions evoked by products in the marketing researches is field which growing rapidly. This "boom" in emotions in relation to consumer researches a partially based on recent findings, that product-evoked emotions can provide useful and actionable data beyond standard sensory and liking judgments can (Desmet & Schifferstein, 2008; King & Meiselman, 2010; Ng et al., 2013a, 2013b). This growing focus on the role of emotions in consumer researches led to the development of new ways into measuring emotions and recruiting already useful methods into this new field. Nowadays, there are number of self-report of product-evoked emotions (mostly in English language), and number of methods to catch facial and vocal behavior and physiology related to emotions, as well as pupil dilation (Cardello & Jaeger, 2016).

Although, self-reports are not considered as the most "objective" measurement of product-evoked emotions, it is still the easiest way to measure emotions in participants and therefore most popular in consumer studies. This is true because of difficulties in measuring and synchronizing response-behavior for further analysis, as well as requirement of hard-wired sensors or artificial positioning of the body and indirectness of interpretation (Grühn & Sharifian, 2016). We won't do comprehensive review of self-report measurement in consumer research literature, interested readers referred to appropriate literature (Grühn & Neika, 2016; Cardello & Jaeger, 2016), and review of methodological issues in product emotion-evoked measurement using self-reports (Jaeger & Cardello, 2016).

As it has been mentioned earlier (Historical sketch), expression of emotions evolved with human phylogenies, and some emotions can me reliably transmitted and perceived through different nonverbal channels (for example, face or voice). Moreover, expressions of emotion could be reliable information about internal feeling, especially when there is no need to control facial behavior. Expressions of emotions in the face seems to be universal (at least for some of them), therefore they are expressed in the same way in

different parts of the world and can signal internal states regardless of sender's culture, age, sex and ethnicity. The most studied channel – face – could give insight into someone's internal world through analysis of contracted muscle and comparing it with prototypical (or major variants or subtle) configurations to understand if that is an emotion (Ekman, Friesen & Hager, 2002b; Matsumoto & Hwang, 2014). Considering widespread globalization, continual reduction of prices on traveling, employee migration and outsourcing, "reading" of nonverbal behavior can be accompanied with less mistakes, if research methods rely on universal versus culture related behavior, and facial expression is first candidate for.

There are only few studies recruiting facial behavior measuring approach in marketing research. The most recent to my knowledge was conducted by Matsumoto and colleagues (Matsumoto, Hwang, Harrington, Olsen, & King, 2011). They've analyzed facial behavior during 30 sec periods after consumers engaged with the product and were interviewed about it using EMFACS (Ekman & Friesen, 1982). They have measured facial behavior and analyzed it to correspondence to emotion prototypes. Their study showed that participants rarely showed expression of positive emotion. Enjoyment was shown only by 3.6% of participants, while the most popular expressions was disgust (28.18%), then non-enjoyments smiles (26.55%), anger (15.50%), contempt (11.35%), fear (8.09%), surprise (6.76%) and finally sadness (4.79%). The most intriguing result was that in 61% of cases, facial expression was incongruent with self-reported emotion. This results once again raising question over self-report's validity.

Two more studies available where facial measurement behavior was used, however in one of them (Derbaix, 1995) coders were trained to identify how many times facial behavior corresponded to prototypical expressions of emotions and were not trained to use any specific coding system. In another study done by Zeinstra with colleagues (Zeinstra, Koelen, Kok & de Graaf, 2009) there was only one coder, that is why it was not possible to measure interrater reliability for the coding and they have used in their study not empirically proven criteria for interpreting facial behavior (Matsumoto et al., 2011).

As mentioned above, only three studies applied behavior measurement approach and only two of them used precise coding of face. One FACS, another EMFACS - abbreviated version of FACS which codes only AUs related to emotions and with restriction of number of reviewing video and in real time speed. It is less time-consuming, however with lower interrater agreement (Rosenberg, personal communication 2014).

1.3.2. Emotions and deception

Because facial expression took the most attention of researches, it's not surprising that facial expression was studied as useful source of deception as well.

Deception leakage is possible because of dual system of facial muscle activation. Face contains 20 striped muscles, the only muscles connected to the skin one the one side and to the bone on another one (Matsumoto & Hwang, 2016). The innervating of facial expressions could be done through two different neural pathways, pyramidal and extra pyramidal tract (Rinn, 1984). First one starts in motor cortex of cerebrum and activates voluntary facial muscles movements, while another one starts in subcortical areas and activates involuntary, spontaneous facial movements, such as emotions. Because both can be activated simultaneously, between those two pathways could happen "tug-of-war" (Matsumoto, Frank, & Hwang, 2013). For example, when a person experiencing joy, one might want to control his or her expression because it is unappropriated to express that emotion in some contexts. For example, in funerals it is not appropriate to smile. In this situation, both pyramidal and extrapyramidal tract are involved; last one contracting facial muscle because of emotion, second one relaxing the same muscle or even involving other muscles to counteract. In the light of example above, through extrapyramidal tract could be contracted m. Zygomatic Major and m. Orbicularis Oculi, Pars Orbitalis, which raise corner lips obliquely and raising cheeks, respectively. That is - prototypical expression of joy (Frank & Ekman, 1993). In contrast, due to display rules, when and where emotions can or cannot be shown, person could make attempt to relax those muscles, or even would contract additional muscles to counteract them, for example m. Triangularis, which lowering lip corners. Another scenario is possible as well. Person might mask emotion of joy with another expression, more appropriate to funeral context – sadness or grief. In this case, both tracts would be activated and innervate muscles concurrently.

This simultaneous "shot" could be successful, and expression of experienced emotion won't show on the face at all. But another consequence possible as well, when person shows "leakage" of truly felt emotion – micro expression.

Micro expression – full-faced or fragmental expression of emotion in the face, which lasts less than ½ second (Frank & Svietaeva, 2014). Difference between micro and macro expression is based on duration of expression. Typically, macro expressions last between ½ to 4-5 seconds (Ekman & Rosenberg, 2005), while micro expressions last less than ½. Another possibility, that expression was not totally attenuated and still was shown in the face but on very low intensity and called "subtle expression". Subtle expression – are expressions of low intensity, A-C (5-point scale) in FACS intensity score, while "strong"

expression refers to D-E intensity (Matsumoto & Hwang, 2014). Several studies shown that training in recognizing both micro expressions (f.e. Frank & Hurley, 2014)) and subtle expressions (Warren, Schertler & Bull, 2009) are related to improvement in deception detection.

1.4. Conclusion

Facial expression is a major source about people's internal state and other characteristics. We can use facial behavior to understand what emotions people feel, how they react to our words, understand when someone is lying. We pay attention to the face. We use photos of our face as avatars in social networks, not other parts of the body. Beyond that, facial expression could be a huge source of information about our intentions, even aggressive. Understanding of facial expression can be used in different domains, personal and business. For proper use of information drawn from the face, we must measure it properly and reliably. As have been mentioned above, there are two broad categories how we get inference from facial behavior. Those two are objective and subjective measuring the face. The former requires expensive equipment (EMG) and is highly time-consuming (behavior measurement approach) and that is why it was rarely used in business context. In this domain, judgment studies (subjective measurement) are much easier to conduct. They require only panel of observers and behavior to judge itself. However, there are caveats as well. Because of it's easier applicability, rest of my thesis would be focused primarily on observer-based approach. How can we draw conclusions about experienced emotion in others by measuring facial expression using people's perception? How can we do it most accurately? To answer those questions, in the next section we will review state of the art of judgment studies and outline pros and cons of different methods in there. In other words, how, different aspects of procedure in emotion judgement studies influence on results or accuracy.

2. Theory

Influence of context of recognition on agreement was studied mainly in emotion judgement studies. Thus, all pros and cons of different methodological differences comes mainly from that literature. Most of them have used very similar design which was even called "standard method". Standard method defined as studies that include "…at least four of the hypothesized basic emotions, posed facial expressions, a forced-choice response format, and a within-subjects design (p. 8, Nelson & Russell, 2013)".

Some of the features of standard method are considered to artificially increase, or push, observers' agreement. These methodological flaws and some other would be reviewed here. The full list presented below (Frank & Stennett, 2001; Russell, 1994):

1. Facial stimuli
    a. Posed expressions
    b. Preselected stimuli
2. Randomization
3. Presentation of multiple faces
    a. Quantity of stimuli
    b. Stimuli preview
    c. Within-subject design
4. Response format

Those flaws would be discussed in great details below. However, before proceeding few notes should be done. First, hereafter, all discussion of emotion recognition (ER) studies would refer only to studies dealing with facial expression. Second, main goal of this thesis is not to undermine on methodological grounds most of studies conducted so far (some of them are mentioned in Historical Sketch) which showed universal recognition, but to understand how we can increase validity for future studies. Nowadays, proves for universality of facial expression is far more solid than showing only universal recognition around the globe (Matsumoto et al., 2008), which could be achieved by other ways then innateness (Ekman, 1973).

First subsection dealing with facial stimuli. Mainly, it is related to concerns with stimuli sincerity (spontaneous vs. posed) as well as with question of preselection stimuli for emotion judgement studies.

2.1. Facial Stimuli

2.1.1. Posed expressions

While studying how people can recognize emotional expression, real spontaneous

expressions were rarely used. There were conducted only 10[1] studies using spontaneous facial expressions in period between 1982 and 2017 years. Here I review them briefly. All three studies before 1982 were criticized based on methodological ground (for more details see Ekman et al, 2013).

In a series of studies by Wagner and colleagues (Wagner, 1990; Wagner, Lewis, Ramsay, Krediet, 1992; Wagner, MacDonald & Manstead,1986) they used emotion-laden slide to evoke emotions in encoders, whose facial behavior should decode other participants.

In Wagner et al. (1986) the mean agreement was 22.9%, low but still significantly higher than chance. Results broken down by emotions, showed that anger, disgust and happiness were all recognized at levels better than chance. Other emotions – fear, neutral, sadness and surprise were not.

In another study, Wagner (1990) used slides that should have evoke 8 emotions - amusement, anger, disgust, happiness, peacefulness, puzzleness, sadness and tender. Those terms were chosen based on free-response labeling by separate group of observers. Overall result was 18.6%, which exceeded chance level. Only expressions of amusement (36.7%) and puzzleness (35.6%) were consistently recognized higher than chance, while peaceful expression was recognized 31.6% times overall, but male decoders never achieved statistical significance neither for expression of male nor female encoders. Other emotions were not recognized higher than chance.

In his third study, Wagner et al. (1992) showed the very same slides to other encoders and new group of observers interpreted their behavior. They have not reported statistic precisely but mentioned that overall agreement was low. Only five out of 32 tests (8 emotions X 2 encoders genders X 2 decoder genders) were statistically higher than chance.

Study by Motley & Camden (1988) differs in the way that they used facial expression evoked during interaction between encoders and research confederates. They have designed situation which should elicit six emotions – anger, confusion, disgust, happiness, sadness, and surprise. In their study, one week later after initial session, encoders came back and posed expression for the same emotions. Twenty observers interpreted those expression.

Mean accuracy for spontaneous expressions was higher than chance. However, when results for highest recognized emotion – happiness – were excluded from analysis, the

[1] Because there is still debates when facial expression emerges in development, two studies where infants' facial behavior was used are emitted (Camras et al. 2006; Yik et al. 1998);

five remaining emotions were not recognized at rates higher than chance. Observers could identify significantly more accurate posed expressions than spontaneous (19.55/24 vs. 6.25/24).

The following six studies share one common feature. Expressions which was used as stimuli were chosen based on FACS coding and corresponding to prototypical expression previously theoretically or empirically related to those emotions.

In study by Hess & Blairy, (2001) 16 facial expressions were used, four examples for anger, disgust, happiness, and sadness. They were presented as 15-sec videos. Videos were taken, when encoders imagined or thought about emotion evoking situation. Only if encoders reported feeling emotion and if expression fit prototype it was used. Decoding accuracy varied between 43% and 87%; specifically, anger – 45%, disgust – 43%, happiness – 87%, and sadness – 75%.

Study by Matsumoto et al. (Matsumoto et al., 2009) differs in one important way. It is the one of two cross-cultural studies where spontaneous expressions were used, and it is the only study where encoders were from different cultures. All facial stimuli were taken during match completion and medal ceremonies in 2004 at Athens Olympic Games judo competition. All expressions were coded using FACS and based on that coding correct labels were predicted. Decoders from England, Japan, US-born and US-international participated in the study. There were examples of facial expressions for only single emotion – anger, contempt, disgust, fear, happiness, sad, or their blends. Here I report results for single emotions only. Average agreement for both match completion and medal ceremonies for all groups together fall into 17.5% to 61.0% range with mean agreement 56.9%[2]. More precisely, anger – 31.2%, contempt – 34.6%, disgust – 24.1%, fear – 17.5%, happiness – 61.1%, sadness – 45.1%.

Next four studies used expressions pulled from the same stimuli set (Ekman, 1980). Every expression occurred naturally in the Papua New Guineans' home environment. Each expression was labeled based on Paul Ekman's knowledge of expression context and FACS coding.

In study by Naab & Russell (2007) observers judged 20 facial stimuli, each depicting one of six basic emotion - anger, embarrassment, happiness, interest, sadness, and surprise; four portrayed one of nonemotional states – perplexed, hesitant and relaxed, the final four encoders portrayed blends of two emotions or emotion and nonemotional state.

[2] There is inconsistence in their results, also mentioned by Kayyal & Russell, 2013. Results in Table 1 (pp. 218–221 in their article) does not fit those in Table 3 (p. 227). When those in Table 1 weighted for number of stimuli, they are still lower than results in Table 3.

Here I present results only for single emotions. Mean agreement for them was 24.3%, which is statistically higher than chance. For each emotion, agreement varied from 7.3% to 45.8%, and only for 4 out of six, predicted label was also the modal response. Anger – 20.8%, embarrassment – 7.3, happiness – 33.3%, interest – 15.7%, sadness – 45.8%, and surprise – 40.6%.

In the following study (Kayyal & Russell, 2013) participants from USA and Palestine judged very same twenty expressions used in study by Naab & Russell, plus one example for disgust and perplexed expressions each.

Average agreement for all three groups by emotion are following: anger – 65%, disgust – 87%, embarrassment - 33%, happiness – 69%, Interest – 48%, sadness – 80%, surprise – 58%. Mean agreement for all emotions – 57% (69% for traditional basic emotions, and 43% for interest and embarrassment together).

In the following two studies reported in Crivelli, Russell, Jarillo & Fernández-Dols (2016) participants were adolescences from Kaduwaga and Vakuta villages in Papua New Guinea. Thus, encoders and decoders were from Papua New Guinea.

In the first study, observers judged facial 5 expression, one example for each – anger, disgust, happiness, sadness and surprise. Participants were not provided with response options, but instead should have provide their own or answer "I don't know" in case they have no such.

Generally, response accuracy was low. None of participants labeled anger expression as anger. The same result emerged for surprise expression. Disgust was labeled correctly only in 6%, Happiness – 13% and sadness – 16%. There was never the case that predicted label was modal

In the second study, adolescents judged same five expressions, but this time they were provided with nine response options – anger, disgust, happiness, sadness, surprise – five terms predicted by Ekman for each stimulus, two most frequently used terms from the first study – "smiling/laughing" and "feels like avoiding social interaction", and two additional options "other" and "I don't know", which were used to eliminate forcing effect of response format.

Only two emotions were recognized at higher than chance level – disgust and sadness and both were the modal response. Anger was correctly recognized only in 13% cases, disgust – 38%, happiness – 17%, sadness – 29%, and surprise – 21%.

Results of those studies drastically different from results obtained using posed facial expression, where mean agreement lie in range between 70 and 90 percent (e.g. see Biehl, Matsumoto, Ekman, Hearn, Heider, Kudoh & Ton, 1997). In review by Nelson and

Russell (2013), for example, mean agreement was 70.9%.

However, methodology of those studies raises some questions. For example, there is evidence, that people can experience emotions and still do not show visible signs of emotions. Studies using EMG (electromyography) showed, that there could be slight muscle contraction, which does not produce visible appearance changes on the face (Ekman, Schwartz, & Friesen, 1978). Two studies (Wagner, 1990; Wagner et al, 1992) have not measure facial expression at all, assuming it was there and relied solely on subjective experience of encoders. Thus, the criteria for accuracy was self-reports. But there is a question was there any sign of emotion in the face.

In study by Wagner and colleagues (Wagner et al., 1986), two naïve judges rated to which extent facial movements were present in the face. Ratings for all emotions were below midpoint on physical movements scale. Moreover, physical movements scale was not correlated with decoders' accuracy.

In another study Motley and Camden (1988) tested how equally responses were distributed across six options. They assumed if that is the case, people are just guessing and can't recognize emotions correctly when they are shown in the face. But they did not. As mentioned above, happiness was the only one emotion recognized higher than chance in theirs study, other five – not. But really this means that there was any sign of other emotions? No. People could use positive-negative distinction and all expression other than positive were randomly allocated to five options because there no facial behavior and they were not positive.

In fact, the only one study where both subjective experience and actual facial behavior were used as selection criteria for stimuli was Hess and Blairy's (2001) study. And this study got the highest recognition rate among others – mean agreement was 62.5%.

For some emotions in those studies there is no proofs they have distinctive emotional expression at all (f.e. peacefulness). It raises additional question of how people could have recognized it, when there is no distinctive sign for that emotional state?

Most emotion recognition studies to date have used posed expression as a stimulus set. There are several rationalizations for that. First, it is highly difficult to get standardized stimuli set of genuinely experienced emotions. People control their facial behavior when they are with other people via display rules, so they conceal inappropriate for context emotions (Ekman, 1972). Furthermore, Fridlund et al. (1990) proposed that people can hide their true expression even when they are alone because of internalized others. Considering that it is almost impossible to capture fully uncontrolled spontaneous facial expression.

Second, emotion signaling function is only one of many which face does. Matsumoto et al. (2009) proposed that because muscles related to emotions can concurrently be recruited in other facial behavior, such as talking, illustration speech or emblematic information, those non-emotional behavior can affect signal clarity of emotion. Reduction in signal clarity in turn can result in lower agreement.

To avoid that, researchers used posed expressions which involve only critical facial muscles theoretically of empirically associated with emotions (Ekman 1993; Ekman & Friesen 1975; Ekman et al., 2002b). However, in some studies stimuli selection criteria is only high recognition by panel of observers (e.g. Elfenbein, Beaupre´, Levesque, & Hess, 2007).

If the posed facial expression does not resemble spontaneous emotion expression there is no other reasons why people would have high agreement in labeling those expressions similarly all around the globe. As you have read previously in Historical Sketch the opposite is true. Summarizing those results, Ekman and colleagues (2013) writes:

"The finding of cross-cultural similarities can help clarify the relationship discussed earlier of posed behavior to spontaneous behavior (see Chapters II, VII, XV). If, as Landis, and later Hunt, argued, posed behavior is a conventional language—socially learned and unrelated to real emotion—then it would be logical to expect, as they did, that poses would be judged differently across cultures. The fact that posed facial behavior was similarly judged across cultures, and that not only were Western poses understood by New Guineans, but New Guinea poses were understood by Westerners, requires either that these conventionalized facial behaviors were, inexplicably, learned the same way in all 14 cultures, or that Landis and Hunt were wrong, and that posed facial behavior resembles and grows out of spontaneous facial behavior. Our view is that posed facial behavior is similar to, if perhaps an exaggeration of, those spontaneous facial behaviors which are shown when the display rules to deintensify or mask emotion are not applied (see page 106). Posed behavior is thus an approximation of the facial behavior which spontaneously occurs when people are making little attempt to manage the facial appearance associated with intense emotion. (pp 167, Ekman, Friesen, & Ellsworth, 2013)"

2.1.2. Preselected stimuli

As mentioned above, there are two different criteria for inclusion expression in final set of stimuli, it could be selected based on high agreement of observers or based on contracted muscles. First way includes other variabilities in study design. For example, in-group advantage (Elfenbein & Ambady, 2002a, 2002b, 2003a, 2003b) was found only

using first selection, but not when facial expression was equal in terms of contracted muscles (Matsumoto, Olide, & Willingham, 2009).

2.2. Randomization

Earlier studies characterized by lack of proper randomization (Russell, 1994). Taking into consideration relative thesis – that is that participants interpret facial expression relative to others (Russell, 1991) randomization could have high influence on interpretation of stimuli (see 2.3.3. subsection for more details).

Nowadays, researchers can use online surveys services, with complex survey flow settings, thus proper randomization should not be a problem anymore.

2.3. Presentation of multiple faces

2.3.1. Quantity of stimuli

How does number of stimuli in study influences on labeling agreement? In their review, Nelson & Russell (2013) mentioned that observers who had viewed fewer faces (6-24) had lower agreement, then those who had viewed more faces (30+), 67% and 78%, respectively. Up to date, there is only one study directly measuring how number of stimuli influences on ERA (Romashov & Shakhraichuk, 2017).

In their study, each participant judged 48 expressions (6 examples x 8 expressions) grouped into 3 blocks. Expressions within each block as well as blocks were randomized. Thus, each stimulus was presented in 1-16, 17-32 or 33-48 place in study, depending on condition.

Decoders randomly were allocated into six conditions, differencing only in order of Block presentation. As expected, depending on presentation order of stimuli in study, the different agreement was. Both, fear and neutral expressions had higher recognition rates when presented in 17-32 and 33-48 places compared to 1-16. Happiness were more readily recognized in 33-48 places compare to 1-16. There was no significant improvement between 17-32 and 33-48 places for any emotion.

Thus, there is data suggesting that number of stimuli can have influence on outcome of the study, even if this influence limited only to some emotions.

2.3.2. Stimuli preview

In some studies, before experimental part, participants review the whole range of expressions for familiarization with stimuli. There is no study which compare agreement directly with or without preview of stimuli. However, familiarization can lead to both, highlighting similarities within expression type and differences between them (Russell, 1994). When participants further presented with list of options, they can draw conclusion

which type of expression is related to each option. This linkage in turn can increase likelihood of applying elimination strategy during experiment. However, as mentioned above there is no direct study comparing that.

2.3.3. Within-subject design

Within-subject design creates the same problem as stimuli preview. Participants exposure to more than one expression, which can artificially augment participants focus on similarities and differences within or between expression type. In within-subject design problem arise because "…all faces but the first are preceded by one or more other faces typically within a short time" (p. 1063; Yik, Widen & Russell, 2013). Several studies shown that depending on the previously seen faces, participants can label same faces differently. This effect called "relative thesis" (Russell, 1991c). Combining within-subject design and with some response formats, can create conditions, where participants can use elimination strategy and tend not to use less appropriate terms when judging expressions, and thus artificially raise agreement level (see 2.4. subsection). Beyond elimination, there is consistent finding of relative interpretation of facial expression.

Relativity thesis can be traced up to study by Russell and Fehr (1987). They have found that same expression can be labeled differently depending on previously seen expression. In their first two experiments, when participants saw neutral expression which was preceded by one anchor expression randomly selected from 19 different emotional expressions (e.g. anger, excitement, boredom etc.) in Experiment 1 or by sadness or happiness in Experiment 2 was indeed labeled differently. For example, in Experiment 2, neutral expression was called sadness after happy anchor in 58% of cases and after sad anchor the modal response for the same neutral expressions was happiness (32%) and second most frequent was surprise (27%). In their third experiment participants rated neutral expression after pairs of emotional expressions used as anchor. The same result emerged. Labeling of neutral expression varied with anchor pair. In the first three experiments they used solely neutral expression as target one.

In fourth experiment, instead, they used five target expressions differed mainly in arousal degree (see Circumplex Model; Russell, 1980). The expressions were: sleepy, semi-sleepy, neutral, alert and aroused preceded by sad or happy anchors. After happy anchor, participants tended to use negative categories (fear, anger, disgust or sadness) while after sad anchor more positive (excitement, happiness, calmness) to describe expression.

In the next two experiments, they have used emotional expression for target stimuli: surprise in Experiment 5 and anger in Experiment 6. In experiment 5, surprise was

preceded by either excitement or fear. It was predicted that modal response would be surprise, but different second most popular expressions were expected.

Predictions based on Circumplex Model were partially confirmed. As expected, magnitude of surprise label varied with condition and was highest in control condition (participants saw only surprise expression). Excitement, but not fear, varied as well, and as expected was lowest after excited anchor.

In Experiment 6, anger expression was proceeded by one of eight anchors - excitement, surprise, fear, anger, disgust, sadness, calmness and contentment or two pairs of them (fear-surprise or calmness-sadness). Categorical ratings for fear and disgust varied depended on condition. Target expression was judged as less afraid after fear anchor, and less disgusted after disgust anchor. Anger was the modal response in most of conditions, but also varied in both categorical and ratings. More interesting is that anger expression was labeled as sadness in two conditions.

Experiments by Russell and Fehr (1987) are interesting, but still controversial. In their review of those results, Ekman and O'Sullivan (1988) highlights the fact that when it was possible, participants labeled expression as calm, which is most close to neutral state among other options. For example, in Experiment 1, calmness was the modal response in 16 out of 19 conditions. The same result emerged for Experiment 3, where calmness was modal response in control and 3 out of 4 experimental conditions. Further, in Experiment 5 and 6, when target expression was emotional and label for it was provided, the correct response was the modal one in all cases. Only Second most popular options varied (for more details see Ekman & O'Sullivan, 1988).

In the following study, Russell (1991c) obtained similar results when correct option was provided as well. In Experiment 1, facial expression of contempt was proceeded either by disgust or sad anchor or was shown alone in control condition. Russell found two critical results. First, unexpected finding was that modal response for control condition was disgust, but not contempt. In other words, for contempt expression participants used disgust label most frequently. Second, as expected, disgust was most chosen option in sad-anchor condition. In disgust anchor condition, modal response was sadness. Thus, in all conditions, contempt was never modal response and anchor expression indeed can influence on following expression.

Second experiment was identical to control condition, with one exception. There was seven remaining expression of contempt from JACFEE (Matsumoto & Ekman, 1988). He showed only 1 expressions randomly selected among them and ask participants to choose one option as well as rate degree to which face expressed all 7 emotions using 4-point

scale (1-4). The same result emerged. Disgust was the highest rated emotion (2.57) as well the modal response in categorical judgement (40%), while contempt was second highest rated (2.33) and categorically judge emotion (29%).

How could be, that expression previously found to be signaling contempt with high agreement (Ekman & Friesen 1986; Ekman & Heider, 1988) was labeled disgust? In the last experiment, participants were randomly allocated into two conditions. In control condition they had judge only target expression – contempt. In experimental condition target expression was preceded by 6 anchors – one example of prototypical expression of anger, disgust, fear, happiness, sadness and surprise. Results for control group was the same as in the previous two experiments – modal response was disgust (60%), the second most frequent was contempt (32%). Absolutely different pattern emerged in experimental condition, where contempt obtained modal response (68%), the second most frequent category was disgust (20%). Thus, Russell (1991c) provided additional support for relative thesis as well as showed pure difference in agreement when within-subject and between-subject design are applied.

In their comment on that study, Ekman, O'Sullivan and Matsumoto (1991a) provides reanalysis of Matsumoto's (1990) study. Although, in his study within-subject design was used, it shows that contempt was modal response for contempt expression in all instances regardless of preceding expression. Ekman and colleagues (Ekman et al., 1991a) proposed that judging multiple faces is more true-to life condition but did not provided any support for that claim (see Russell, 1991b), but still admitted that participants could have used elimination strategy.

In the following study, Russell (1991a) found that when eight expressions from JACFEE set (Matsumoto & Ekman, 1988) studied using between-subject design and free-response format they still were most frequently labeled as disgust (10%). Contempt was mentioned by 2% of participants. In second study, the same facial expressions were judged using rating scales. Participants should have rate to which each of emotions is presented in the face – contempt, anger, boredom, disgust, frustration, and scorn (options were not contempt-related in study 1 and close to target expression according to circumplex model; Russell, 1980) and to which degree, using 4-point scale. Once again, contempt wasn't the most intense emotion, it was fourth after boredom, disgust and frustration. For more detailed review of that study see Ekman, O'Sullivan and Matsumoto (1991b)

Pochedly and colleagues (Pochedly, Widen & Russell, 2012) extended our understanding of relative thesis. They reported two studies with main assumption that

"nose scrunch face" (nose wrinkled, and upper lip raised) could be a sign of disgust, only if it preceded by "right" anchor stimuli. In the first study participants were randomly allocated into three conditions: (1) anger scowl, (2) no face, or (3) sick face, depending on anchor stimuli – face preceded the target nose scrunch face. There was fixed order: happiness, surprise, fear, ANCHOR, and target stimuli – scrunch face. Sick face was posed especially for this study: FACS AUs 6 + 10 + 25 (cheeks raised, upper lip raised, lips parted). Participants (children) could judge expressions using those options: angry, disgusted, embarrassed, happy, sad, scared, surprised. Once again, confirming relative thesis, how nose scrunch face was labeled depended on the face preceded it. When anchor was anger, children were most likely to call scrunch face as disgusted, while in sick face condition the least. Pairwise comparison showed that in "anger scowl" condition, children significantly more often called nose crunch face as disgust, than in "no face" or "sick face" condition, with no difference for the last two.

Second study was almost the same, with few exceptions. This time, participants were adults. There were two anchor order condition: in the first one on the second position (happiness, ANCHOR, surprise, fear, target face – nose scrunch face) and in the second on the forth (happiness, surprise, fear, ANCHOR, target stimuli – nose scrunch face), and there was no "no face" condition.

Second study replicated results of the first one. When there was anger scowl anchor presented, target face was considered to be disgust, but when anchor was "sick face" modal response was anger. Difference was significant. Regardless of anchor position – second or forth, or poser condition – multiple or single, effect of anchor face was replicated.

However, there are some critical notions can be made. First, it is still the question why second condition in the first study was called "no face" and was considered as no anchor face if there was one; instead of anger scowl or sick face, previous face was "fear gasp". Second, both "nose scrunch face" stimuli were not resembling of disgust mentioned earlier in their study – nose wrinkling and upper lip raising. Both expressions at least on half were consisted of action units that could not be considered as a part of disgust. FACS AU code for "crunched nose face" were 4 + 7 + 9 + 25 and 7 + 9 + 18 + 23, pulled from POFA (Ekman & Friesen, 1976) and posed for those studies expression, respectively. According to Ekman et al (2002b, Table 10-1, p174) they referred to, none of expressions could be considered pure examples of disgust. The only one AU among that expressions – AU 9 – could be considered as a component of disgust expression. Furthermore, first expression includes AUs 4, 7 and second 7, 23, AUs all are components of anger

expression, but not disgust.

In another study (Yik, Widen & Russell, 2013), participants from USA and China were allocated in three conditions. The difference between conditions were in terms of anchor expression which they seen previously target expression of disgust - anger, sad and sick face. Filler expressions were happy, fear and surprise expressions. Participants saw expressions in the following order: ANCHOR, happy, fear, surprise, ANCHOR, fear, happy, ANCHOR, and target expression – disgust; ANCHOR differenced depended on condition. As in previous study, agreement on disgust varied with condition, with the highest after anger anchor and the lowest after "sick face", whereas sad expression was in the middle. Mean agreement for three conditions was 68%, 47% and 24% for anger, sad and sick expression, respectively.

Finally, in recent study by Romashov (2018) absolute agreement was compared for between and within-subject design. Each emotion was compared in three conditions: within-subject design (always last expression), between-subject design (always first expression) as well as in control condition, where order of stimuli was randomized. Half of the expressions got statistically lower agreement in between-subject design compared to within-subject design: anger (36.4% vs 65.5% for between and within-subject design, respectively), fear (40.5% vs 67.6%), happiness (58.8% vs 100.0%) and neutral expression (71.4% vs 97.4%). However, contempt and surprise got similar agreement in all three conditions. Interestingly, not all expressions were the least recognized in between-subject design. Control condition got the lowest agreement for contempt, disgust and sadness, and in later two statistically lower than in within-subject design.

Well, could the lower agreement be only due to unfamiliarity with procedure? However, how absolute agreement for some emotions could be high in between-subject design (when expression presented in the first place) and low for others if the problem is in unfamiliarity with the procedure? In Romashov's study (2018), for example, highest agreement in between-subject condition was for sadness (90.3%) and lowest for anger (36.4%). If the participants' low agreement can be justified by unfamiliarity with the procedure, how can sadness get so high result?

2.4. Response Format

Among all methodological flaws of judgement studies, the response format is sought to be the biggest one (Frank & Stennett, 2001). Most studies to day used forced choice format (Nelson & Russell, 2013). Recently, fixed choice becoming more popular like a solution for forced choice drawbacks (reviewed below). There are some more approaches to collect responses, like free labeling or rating scales, but they are applied less frequently

and share some of forced choice flaws (Wagner, 1997), thus, we will focus on the first two – Forced and Fixed Choice formats.

2.4.1. Forced choice

In forced choice method, subjects are restricted to a specified list of alternative labels and are forced to choose one of the labels, while fixed choice method is very similar to forced choice, but subjects are provided with a category such as ‘‘none of these’’ as well (Wagner, 2000).

Series of studies by Russell and colleagues (DiGirolamo & Russell, 2017; Russell, 1993, 1994) have consistently shown that forced choice method can artificially push agreement or provide high agreement when there is no “correct” response option.

In the first study, Russell (1993) presented to subjects one expression of four basic emotions (anger, contempt, disgust and sadness) and provided them with list options, which did not contain the “correct" one. For example, subjects who saw anger expression, depending on condition could pick answer from two lists. First one contained happiness, surprise, contempt, fear and interest, while second one was almost the same, with exception, contempt was replaced with frustration. Based on structural model (Russell & Bullock, 1986; Russell & Fehr, 1987) he proposed that when target expression was anger, predicted best options would be contempt or frustration; for disgust and sadness predicted option was contempt; for contempt either boredom or disgust. Although none of provided labels were synonym for target emotion, they always were modal response and mean agreement for all conditions ranged from 46.3 to 96.3%. Generally, agreement was comparable to those in previous studies. For example, facial expression often labeled as anger here were called as contempt or frustration with 76.2% and 96.3% agreement, respectively.

Second study was practically the same, with some modifications. Subjects saw few new expressions - anger, fear, sadness, disgust and surprise, there were more options and order of predicted option was different in each condition. Predicted options were as follows: for anger either contempt or disgust; for fear surprise; for sadness fear; for disgust anger; for surprise fear. As in previous experiment, the predicted term always was the modal, and agreement ranged from 70.0% to 93.75%.

DiGirolamo & Russell (2017) proposed that high agreement between participants which facial expression could signal each emotion could be artefact of the method used. Especially, they pointed that participants can use elimination of least applicable term and conducted seven experiments to support this hypothesis. In their first experiment participants saw 5 expressions: fear, happiness, sadness, surprise and wink face (FACS:

R7B + L10D + L46), expression previously never related to any emotion, neither theoretically nor empirically. Order of stimuli presentation was always the same: happiness, sadness, fear, surprise and TARGET – wink face expression. Participants was provided with the following response list: delighted, sorrowful, fearful, TARGET, and astonished. Depending on condition, target response option was: (1) disgusted, (2) annoyed, (3) playful, or (4) mischievous.

The results are compelling. In all conditions, modal response was target one, with mean agreement between 76% and 96%, and all target labels were used significantly more frequently than any other label provided. Thus, method used forced participants to choose predicted label even if it does not match expression. Second major result was tendency to use target response option for wink expression when it hasn't been used for the first four trials more frequently. For example, when participant used disgust in the first four trials, they used it for target expression 46% times, but significantly more frequently when they haven't used it previously (86%). Same results emerged in playful (73% vs. 93%) and mischievous (67% vs. 96%) conditions. Results for annoyed were in predicted direction (60% vs. 88%), but not significant ($p = .09$)

In the following experiment, they used more expressions: wink face, puffed cheeks (34D + 17A) or lip funnel (25D + 22D) and new response option list: sorrowful, delighted, fearful, nonplussed, and astonished, always in that order. Nonplussed was used as predicted target response and because no one ever proposed that it has its own expression. Replicating first experiment, nonplussed always was modal response for all three target expressions (agreement rates were between 82% and 93%) and was chosen significantly higher than any other label. When participants used nonplussed prior to target expression, they tended to use it significantly less frequently for target expression, than when it was not previously used (76% vs. 96%).

There is reversal type of forced choice procedure for studying emotion recognition, where participants are provided with one term or a short story describing emotion-evoking event, and then they should choose one of expression as a response option (Wagner, 1997). Could this procedure be contaminated with elimination process too? Third experiment by DiGirolamo & Russell (2017) was designed to test it. In this experiment, participants matched term to one of expressions from array of five, always the same – happiness, sadness, scared, surprise and wink face. Terms were presented in the same order for all participants and were the following: delighted, sorrowful, fearful, nonplussed, and astonished. For the nonplussed term, 60% of participants chosen wink face, with agreement significantly higher than chance, once again showing forcing of

method.

In the fourth and fifth experiments, participants were randomly allocated to conditions differing only on anchor expression - anger, happy, sad and anger, fear, surprise - seen before target expression – sadness or fear for Experiment 4 and 5, respectively. For example, in forth experiment, all participants viewed stimuli in the following order: happiness, anger, ANCHOR, surprise, ANCHOR, fear, ANCHOR, and sadness. In experimental conditions, ANCHOR was replaced with corresponded expression (for anger first anchor was excluded) and had to pick one word out of happy, sad, angry, surprised, scared, or disgusted options. As predicted, the use of disgust and sad response options varied with condition. For the same expression modal label was sadness in control, anger and disgust condition, while disgust in sad condition. When participants used sad in first trials, they tended significantly less frequently use it for target expression (42% vs. 79%). The same result emerged for disgust (14% vs. 39%). In fifth experiment, the same result emerged. The use of both fear and surprise varied with condition. Surprisingly, in both control and anger condition, target fear expression was labeled as surprise and it was labeled as fear in surprise anchor condition. Finally, in fear anchor condition it was labeled both fear and surprise equally frequently. When participants have used label surprise, they used it 25% once again for target expression, while it was labeled 52% surprise times when option hadn't been used before. For scared label, result was 42% and 54%, respectively, but this difference wasn't statistically significant.

However, all those studies do not resemble "standard method" because of fixed order and low number of stimuli. To remedy that gap, sixth experiment conducted. Participants in this experiment saw in random order four examples of happiness, sadness, fear, surprise, anger, disgust, and target expression – lower lip depressor (AUs: 16+25+26), thus whole range contained 28 photos. But depending on condition, one of non-target expression type was wholly omitted. Participants provided with the following options: happy, surprised, scared, angry, disgusted, and sad.

Results speak for themselves, using of anger, surprise and disgust labels for target expression varied with condition. In all instances, the highest endorsement of term corresponded to condition where it was omitted. For example, when there was lack of anger expressions, the anger was endorsed most frequently for target expression. The same was true for both surprise and disgust conditions. In other words, when there are more expressions (24) and order is randomized, people still use process of elimination.

Seventh experiment continued the line and was designed to investigate would people pick non-existed word in English – Tolen - for expression never proposed to be signal of

any emotion – puffed cheeks, similar but not the same expression used in Experiment 1. In single-prior condition order of stimuli presentation was following: happiness, sadness, anger, and surprise. In double-prior condition, there were two examples of each emotion and they were presented in the following order: happiness, sadness, happiness, anger, sadness, surprise, anger and surprise. Target expression – puffed cheeks – was always in the last place. Response options were: exuberant, melancholy, wrathful, awestruck, and tolen.

Fifty three percent of participants labeled puffed face as tolen, note, non-existed word. Contrary to expectations, in double-prior conditions agreement percentage for tolen was lower than in single-prior condition, but this difference was not significant (37% vs. 68%, respectively). Moreover, tolen was used for expressions of basic emotions by 21% of participants. Once and again, participants who used label tolen for other than puffed cheek expressions, rarely used it once more time for target expression, while those who had unused it significantly more frequently labeled puff cheeks as expression of tolen (13% vs. 63%).

Those experiments showed that people tend to use process of elimination to choose correct response for expression. But, in most of them, there was no correct answer provided, thus it is still questionable to which extend those results resemble studies described in Historical sketch. As a solution, it was proposed, that adding option such as "none of those" would decrease forcing effect.

2.4.2. Fixed Choice

Adding option such "none of those" to standard forced choice format decrease it's forcing characteristics, but in all those studies there were no "correct" response provided. For example, DiGirolamo and Russell (2017) in their report described one more experiment, very same as seventh one with one exception. They have added "none of above" response option and obtained dramatically different result. In this case, 64% of participants labeled – puffed cheeks - target expression as "none of those above", but still there were 28% of participants who labeled it as tolen (vs 57% when "none of those" wasn't provided).

Recent study by Romashov (2018) showed how agreement may vary in Fixed choice format depending on anti-forcing term. In experimental design like Russell's (1993), participants saw 8 expressions type (anger, contempt, disgust, fear, happiness, neutral, sadness and surprise), with contempt always on the last place. Depending on condition, they were provided with different sets of response list. In Forced Choice condition, participants were provided only seven response options: anger, disgust, fear, happiness,

neutral, sadness and surprise, thus with no "correct" option for contempt expression. Three other conditions were called Fixed Choice and varied with additional options: none of those, other or both. Results showed, that participants tended to use nonemotional option in "none of those" (70.2%) and "none of those & other" (75.0%) conditions significantly more often than in Other condition (47.4%) but did not differed between themselves. Moreover, response distribution varied significantly with condition.

When there is correct option provided, would be there any need in using option such as "none of those"? Only few studies conducted so far and all of them a reviewed below.

In series of experiments, Frank and Stennett (2001) compared those two response formats using between-subject design, posed expression and stimuli that was normed by American observers and got 80% or higher agreement.

There were two conditions, in both subject seen one expression from pool of 12 expressions portraying 6 emotions by male and female. All expressions were posed by different people. After seeing the face, participant had to select one of six responses: anger, disgust, fear, happiness, sadness, and surprise, each portrayed in the stimuli. Differences between two conditions were in presence of "none of those terms are correct" next to the other six.

Results shown, that the modal response for every stimulus was predicted by universalist, both in forced and in fixed choice condition as well as for both actors' gender. Moreover, the all expressions were significantly higher than chance and even when 25% threshold were set. The mean agreement for forced choice were 84.8% and 80.9% for fixed choice, which was not significantly different from each other.

The study described above suggest that recognition rate would not decrease in Fixed Choice format when there is "correct" term provided and "none of those terms are correct" option. In the second study, they once again tested if people would use "none of those" option when there is no correct answer. They took one expression for anger, disgust, fear, happiness, sadness, and surprise, half posed by women. With each expression were provided 5 response option with excluding correct response. In Fixed Choice "none of those terms are correct" option was added. For Forced Choice condition the same patterned emerged as in Russell's (1993) study – participants agreed on an incorrect response statistically higher than chance for every emotion but fear, which participants not agreed at level higher than chance for any option. Anger was considered as disgust, which in turn were considered as anger; happy expression as surprise; sad as disgust; and surprise as fear. In Fixed Choice condition, the modal response was "none of those terms are correct" for all emotions, with exception for anger, which still was

considered as disgust; all statistically higher than chance.

In their third experiment, participants judged fear and senseless expression using Forced and Fixed Choice formats. There were two-line drawing stimuli presented, one is facial expression of fear and the second one is senseless expression with wrinkles and bulges pattern that never occurs in human faces. The fear expression was always labeled as fear statistically higher than chance, while modal and the only statistically higher than chance response for senseless expression was disgust in Forced Choice condition, and "none of those terms are correct" in Fixed Choice condition.

Nevertheless, Frank and Stennett's (2001) study confirmed that overall agreement rate was similar for Forced and Fixed Choice not in standard method, where within-subject design used, but in between-subject design. In their studies, each participant judged only one expression and that is why it is highly unlikely they could apply elimination strategy (DiGirolamo & Russell, 2017).

To remedy that gap, Romashov and colleagues (Romashov, Shakhraichuk, & Daniluk, 2016) compared the absolute agreement for 7 emotions and neutral expression in within-subject design. Contrary to Frank and Stennett's (2001), for contempt, neutral, sadness expressions and total results there were significant differences. Thus, forced choice in combination with within-subject design can indeed artificially inflate agreement.

2.4.3. Extended response options

How does extended response list influences on participants' agreement?

In his third study, Russell (1993) presented to participants one of two anger expression with seven response options: anger, determination, frustration, hatred, hostility, jealously, and pain; all terms close to anger according to Russell and Fehr's (1987) model, but only anger response indeed semantically denotes emotion of anger. Although anger response considered to fit best, it captured only 5% and 20% for two stimuli photos. For both expressions, the modal response was frustration (45.0% and 35.0%) and the second most frequent response was determination (40.0% and 23.3%).

In another study by Frank and Stennett (2001) participants had to recognize six basic emotions (anger, disgust, fear, happiness, sadness, and surprise) using both Forced and Fixed Choice format. However, this time response list was enhanced with additional four responses – alarmed, bored, contempt, and excited. All alternatives were chosen once again on basis of Russell and Fehr's (1987) model. In contrast to the previous study, all expressions were labeled as predicted by universalist point of view, both in Fixed and Forced Choice condition. The lowest agreement was 65% for surprise in both conditions and for fear in Fixed Choice condition, but still it was statistically higher than 25%

threshold.

2.5. Conclusions

Short insight presented above clearly shows how agreement could influenced be research design. Every mentioned field – posed versus spontaneous expressions, original versus preselected stimuli, randomization, quantity of stimuli, stimuli review, within- versus between-subject design and different response formats – all can influence on participants' agreement. However, as Frank and Stennett (2001) have mentioned, most of them could push agreement slightly, however, response format in standard method – Forced Choice – is real problem. Proposed solution for Forced Choice – Fixed Choice is not studied sufficiently. In the following section, I will present experiment, which is directly comparing agreement using Unbiased hit rate, both in Fixed and Forced choice formats.

3. Research

In this section I will present methodology, structure of the research, data analysis methods and obtained results in experiment.

3.1. Aim of the experiment

Main aim of current experiment was to provide additional support for the notion, that Fixed Choice could indeed decrease forcing effect in judgement studies. To do so, we have compare emotion recognition rates in (1) *Forced and Fixed response format* using (2) *Unbiased hit rate*.

1. There were only two studies which compared agreement in Forced and Fixed response formats directly (Frank & Stennett, 2001; Romashov, Shakhraichuk, & Daniluk, 2016). In Frank and Stennett's experiment was no difference for total result, and no analysis for separate emotions was presented. Romashov et al. (2016) found difference for contempt, neutral, sadness expressions as well as for total result. Those difference can be explained by differences in experiment designs. In Frank & Stennett (2001) experiment between-subject design was used, where participants saw only one expression. In Romashov et al. (2016) experiment within-subject design was used, where participants judged 64 expressions in a row. The later one is representative of standard method, where participants can apply elimination strategy (DiGirolamo & Russell, 2017).
2. Those two studies used Conventional hit rate, which is sensitive to response and stimulus bias. In the current experiment, we use Unbiased hit rate (Wagner, 1993, 1997) which is insensitive to them.

Fixed Choice was proposed as solution for one drawback of standard method – Forced Choice which is artificially increase agreement. However, it does not account for response bias. Unbiased hit rate does. It considers cases when responses are distributed among other than target expression, but it's can't solve the problem of guessing and when participants select target response only by chance because they forced to. As been described earlier (see Subsection 2.4.1.), participants tend to use elimination strategy (DiGirolamo & Russell, 2017) when presented with standard method – or precisely when Forced Choice and Within-Subject design are combined. When they believe, the correct option for target expression is not presented, in the Forced Choice participants tend to choose one option from the list that fit least to other expressions. In this case, Fixed Choice, where responses such as "none of those" or "other" are presented, participants

are not obligated to choose perceived incorrect option. They choose "none of those" when they sure there is not correct option.

For the purpose of the current experiment, we have reanalyzed data from Romashov et al. (2016) report using Unbiased hit rate. If there still would be difference between two response formats, when response bias and stimuli biases are controlled (for what Unbiased hit rate was designed), there would be more reason to believe that forcing to choose responsible for this difference. In other words. we would have additional proof for usefulness of Fixed Choice format for judgment studies, which is as easy to use as Force Choice, but less forcing for participants.

3.2. Hypothesis

Hypothesis 1: Mean agreement with Unbiased Hit rate would be higher than Probability chance for all expressions in both Fixed and Forced Choice formats. Although, there was no study before directly applying Unbiased hit rate for Fixed Choice, we have no reason to expect that emotion would be recognized lower than Probability chance. In most studies with Fixed Choice and Conventional hit rate, emotions were recognized at higher than chance level (f.e. Biehl, Matsumoto, Ekman, Hearn, Heider, Kudoh, & Ton, 1997; Wingenbach, Ashwin, & Brosnan, 2016).

Hypothesis 2: Mean agreement with Unbiased Hit rate would be lower in Fixed Choice compared to Forced Choice Format. We expect to be lower agreement in Fixed Choice, because it reduces forcing effect of Forced Choice format (f.e. Russell, 1993). Thus, participants who are sure there is no correct answer enlisted in response options, would choose option "none of those" or "other". Moreover, elimination strategy is less applicable in Fixed Choice response format (DiGirolamo & Russell, 2017).

Hypothesis 3: Correlation between Unbiased Hit rate and Conventional Hit rate would be higher in Fixed Choice format compared to Forced Choice format. We expect it for the same reason as in Hypothesis 2.

3.3. Participants

Eighty-two psychology students of Taras Shevchenko National University of Kyiv took part in experiment. They were randomly[3] allocated either in Forced Choice (N = 50, 10 males) or Fixed Choice conditions (N=32; 4 males). Both groups, Forced and Fixed, were similar in age M = 21.36 and M = 20.40, respectively; $p = 0.385$, $t = 0.874$, $df = 80$) or gender proportion ($p = 0.486$, $t = 0.701$, $df = 80$).

[3] Lower number of participants in Fixed Choice format is due to higher number of unfinished questionnaires

3.4. Methods

Unbiased hit rate

Wagner (1993) reviewed methods for accuracy calculation between 1979 and 1991 and draw conclusion that none of them are met the criteria's listed below:

1. To be insensitive to response bias
2. To be insensitive to stimulus bias
3. To be applicable to analyze separately accuracy for each stimuli type
4. To be useful for further comparability between studies with different number of response class

He proposed Unbiased hit rate instead. Unbiased hit rate is method for calculating response distribution in judgement studies (Wagner, 1993, 1997). Using example from Table 1., Unbiased hit rate (H_u) for stimulus/response category 1 can be calculated using following formula:

$$H_u = \frac{a^2}{(a+b+c) \times (a+d+g)}$$

Table 1. Unbiased hit rate calculation example

	Stimulus			
Response	1	2	3	total
1	a	b	c	a+b+c
2	d	e	f	d+e+f
3	g	h	i	g+h+i
Total	a+d+g	b+e+h	c+f+i	N

The difference between Conventional hit rate and Unbiased it rate could be presented on example of Sorenson (1975, 1976) studies, where his participants labelled every expression as anger. For example, if category 1 is anger, 2 is contempt and 3 is disgust and there was equal number for each stimulus category (f.e. N = 10). In Conventional hit we would say that participant recognized anger in 100%, and two other emotions – contempt and disgust in 0% cases. However, does people recognize anger in 100% if they use this label for non-anger expressions? The answer could be obtained with Unbiased hit rate. Formula and results would be the following:

$$H_u = \frac{10^2}{(10+10+10) \times (10+0+0)} = \frac{100}{300} = 0.33$$

Because people tend to use anger response for non-anger expressions, they are not using anger options correctly all the time. That is what crucial in calculating H_u.

But what happens when there is unequal number of stimulus? For example, there are 20 anger stimuli, 10 for contempt and 10 for disgust. Formula and results would be the following:

$$H_u = \frac{20^2}{(20+10+10)\times(20+0+0)} = \frac{400}{800} = 0.50$$

Those two examples show clearly how insensitivity to response bias as well as stimulus bias could lead to better understanding of emotion judgments. Unbiased hit rate is easy to interpret, because it's results range between 0 and 1, easy to compare between studies and there is easy way to compute Probability chance (P_c).

Nowadays, unbiased hit rate is not widely applied, but becoming more popular. In every article it was used, it is presented simultaneously with Conventional hit rate, but never separately. Example of its application can be found in many studies (f.e. Biehl et al, 1997; Wingenbach, Ashwin, & Brosnan, 2016). However, Unbiased hit rate have never been applied with Fixed Choice format. To remedy this gap, we have conducted experiment described below.

Stimuli materials

For our experiment, we used Radbound Faces database (RAFD). RaFD is a set of pictures of 67 models (39 Caucasian males and females, 10 Caucasian children, both boys and girls, and 18 Moroccan Dutch males) displaying 8 emotional expressions – 7 emotions: anger, contempt, disgust, fear, happiness, sadness, surprise; as well as neutral expression – All expressions were posed based on DFAT (Ekman, 2007), with three gaze directions: straight ahead, avert to the left and avert to right (Langner, Dotsch, Bijlstra, Wigboldus, Hawk & van Knippenberg, 2010). For current experiment were used expressions posed only by Caucasian adults, with straight gaze and head position. We took 64 expressions of randomly selected 4 male and 4 female actors from RAFD.

Procedure

Every participant answered short demographic questionnaire (gender; age; country) and then started the experiment. In experiment, participants seen 64 expressions (8 types X 8 actors), one at a time. Order of stimuli were randomized.

In both conditions, they saw identical instruction: "Your task in this experiment is to

look at the facial expressions in each photo and decide what person in the photo feels. Under the photo there will be listed response options. Select one option that best represents the facial expression in the picture".

As mentioned above, participants were randomly allocated either into Forced Choice condition or Fixed Choice condition. The difference was only in response options available for participants. In Forced Choice condition, participants should have chosen one option from 8 following options: anger, contempt, disgust, fear, happiness, neutral, sadness and surprise. In Fixed Choice condition two additional options were available – none of those and other. Both were added for reducing forcing to pick one of presented options. When participants chosen Other response option, they were provided with field where they could type their own label. All response options, except Other (which always was in the last place) were randomized.

Participants completed experiment online individually at home using their personal PC with internet. Reusable link was shared among student. Survey prepared using Qualtrics platform (www.qualtrics.com).

3.5. Results

3.5.1. Hypothesis 1: Mean agreement with Unbiased Hit rate would be higher than chance for all expressions in both Fixed and Forced Choice.

Following the procedure for calculating probability chance proposed by Wagner (1997), we first built confusing matrixes for each participant. Returning to our example (see Table 1 in the 3.4. subsection), probability chance (P_c) for each expression type is calculated using following formula (Wagner, 1993, 1997):

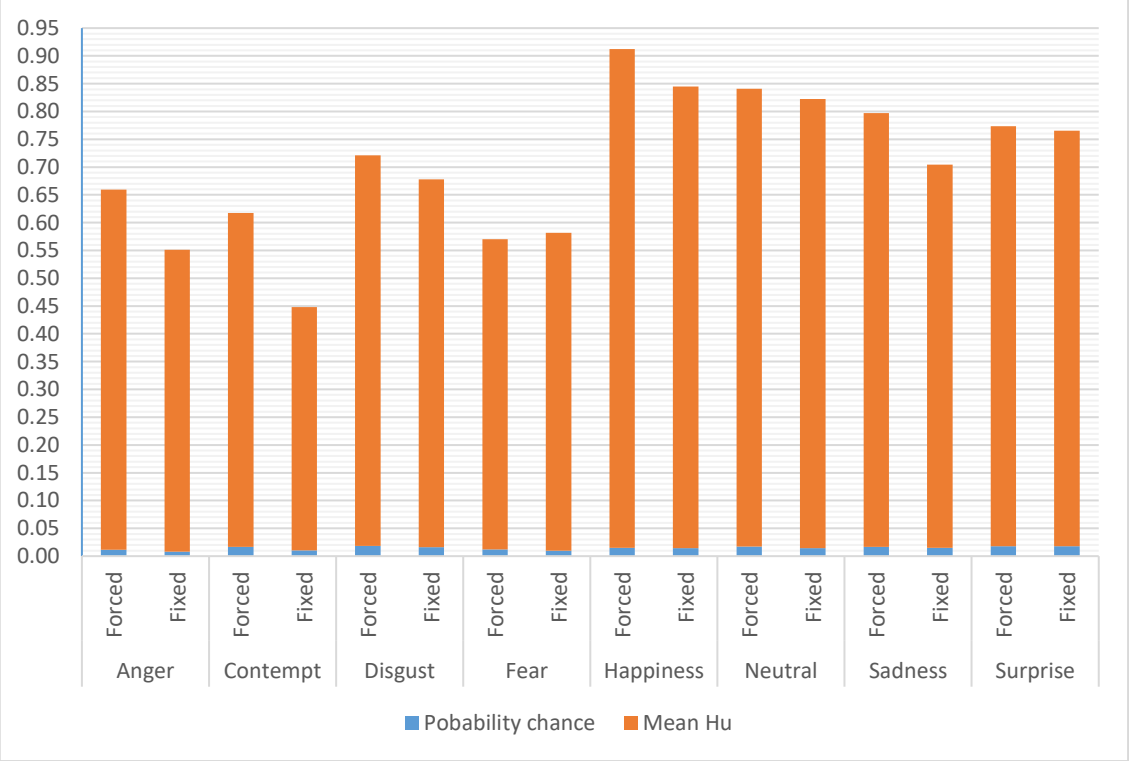
$$P_c = \frac{(a+b+c)}{N} \times \frac{(a+d+g)}{N}$$

Then we calculated mean P_c for each expression type (anger, contempt, disgust, fear, happiness, neutral, sadness and surprise) for two conditions (Fixed and Forced Choice) separately. Mean P_c and H_u are presented on Graph 1 below. Complete data (P_c and H_u) is presented in Annex 1.

Mean Unbiased hit rate (H_u) were compared with probability chance (P_c) using repeated measures *t*-test separately for each expression type in both conditions. Paired-Samples T-Test selection was based on Wagner's (1997) suggested. As you have mentioned on Graph 1, mean H_u was always higher than P_c for each expression in each condition. As expected, there was a significant difference in the scores for Probability

chance and Unbiased hit rate (for all p = < 0.001) in predicted direction. Specifically, these results suggest that participants agreed (H_u) on predicted label significantly higher than would be predicted by chance (P_c). Those results strongly support Hypothesis 1 - Mean agreement calculated with unbiased hit rate is indeed higher than probability chance for all expressions in both Fixed and Forced Choice response formats. Thus, first hypothesis was confirmed.

Graph 1. Probability chance (Pc) and mean agreement (Hu)



Source: own elaboration

3.5.2. Hypothesis 2: Mean agreement with Unbiased Hit rate would be lower in Fixed Choice compared to Forced Choice Format

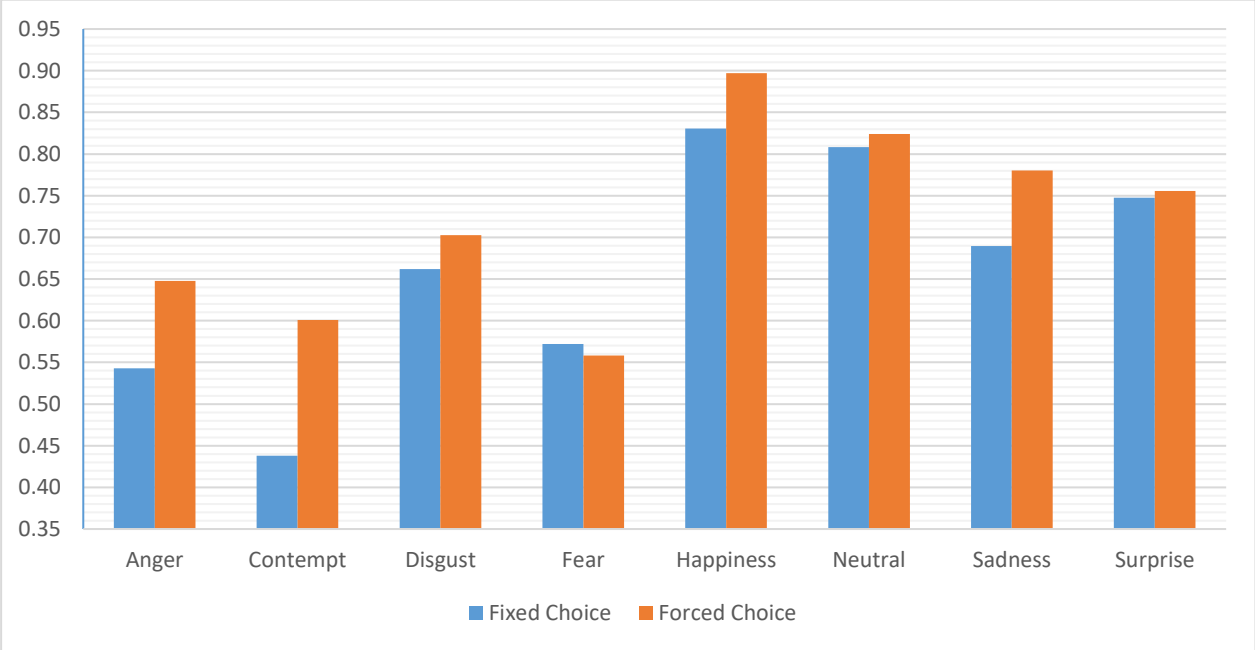
To test second hypothesis, we used Independent-Samples T Test. We compared agreement for each expression type (anger, contempt, disgust, fear, happiness, neutral, sadness and surprise) in two response formats (Fixed vs. Forced choice).

As you can see on Graph 2, mean H_u for anger, contempt, disgust, happiness, neutral expression, sadness and surprise were lower in Fixed Choice, but fear was in opposite direction. However, only contempt (M = 0.44 and M = 0.60, p = < 0.05, t = 2.203, df = 80) and sadness (M = 0.69 and M = 0.78, p = < 0.05, t = 2.094, df = 50) got significantly lower agreement in Fixed Choice format compared to Forced Choice.

Because only contempt and sadness were significantly different, although other emotions (except fear) were in predicted direction, second hypothesis was only partially

confirmed.

Graph 2. Mean agreement (Hu)



Source: own elaboration

3.5.3. Hypothesis 3: Correlation between Unbiased Hit rate and Conventional Hit rate would be higher in Fixed Choice format compared to Forced Choice format

We have used the Fisher r-to-z transformation calculator (http://www.vassarstats.net/rdiff.html; 10.06.18) for comparing correlation between Conventional and Unbiased Hit rate in Fixed and Forced Choice format. Correlations and results of test are presented in Table 2. As you can see, in generally correlation between Unbiased hit rate is higher with Fixed Choice. Only for happiness, correlation between Unbiased hit rate and Forced Choice response format was in opposite direction.

Table 2. Correlation between Conventional and Unbiased hit rate for each emotion.

	Anger	Contempt	Disgust	Fear	Happiness	Neutral	Sadness	Surprise
Fixed Choice	0.98	0.95	0.81	0.98	0.85	0.87	0.89	0.77
Forced Choice	0.96	0.91	0.73	0.93	0.92	0.61	0.82	0.68
p* =	0.27	0.16	0.42	< 0.01	0.15	< 0.01	0.26	0.44

* two-tailed

As expected, for most of emotions (except Happiness), correlation between Conventional and Unbiased hit rate were generally higher in Fixed Choice format compared to Forced Choice format. However, this difference reached statistical significance only for Fear and Neutral expressions (both p = < 0.01). Third hypothesis was partially confirmed.

3.6. Conclusions

Experiment described above have confirmed our hypothesis, fully (Hypothesis 1) and partially (Hypotheses 2 and 3). As expected, agreement for all emotions were significantly higher than chance, irrespective of response format. Current experiment is the first one, where Unbiased hit rate was applied for Fixed Choice response list.

Second hypothesis was partially confirmed. Mean agreement for every emotion (except fear) was lower in Fixed Choice format, thus in predicted direction. However, only for contempt and sadness this difference was significant. This is first time, when Unbiased hit rate was used for comparison of Fixed and Forced Choice format. These results give us additional support for the notion, that Forced Choice indeed artificially pushes agreement. Applying unbiased hit rate gave us possibility to exclude response and stimulus biases typical for Conventional hit rate. Even after that, two emotions (contempt and sadness) had lower agreement in Fixed Choice compared to Forced Choice. Interestingly, contempt and sadness were among those three (contempt, neutral, sadness) expression types which had different agreement in Conventional hit rate as well (Romashov et al., 2016).

Finally, for most of emotions (except Happiness), correlation between Conventional and Unbiased hit rate were generally higher in Fixed Choice format compared to Forced Choice format and this difference was significant for fear and neutral expressions.

Those results once again show that Forced Choice format indeed artificially increases agreement. Future researches should use Fixed Choice, which is as easy to apply as Forced Choice, but which has lesser forcing effect.

Applicability of those results extend far more than just for academic use. If we are interested in precise measurement of emotion judgement where panel of observers decide what does face, body express or product which emotion evoke, it is much more fruitful to use Fixed Choice response format.

Bibliography

Articles and books (alphabetically)

1. Barrett, L. F., Lindquist, K. A., & Gendron, M. (2007). Language as context for the perception of emotion. Trends in cognitive sciences, 11(8), 327-332.
2. Biehl, M., Matsumoto, D., Ekman, P., Hearn, V., Heider, K., Kudoh, T., & Ton, V. (1997). Matsumoto and Ekman's Japanese and Caucasian Facial Expressions of Emotion (JACFEE): Reliability data and cross-national differences. Journal of Nonverbal behavior, 21(1), 3-21.
3. Buck, R., & Miller, M. (2016). Measuring the dynamic stream of display: Spontaneous and intentional facial expression and communication. In D. Matsumoto, H. C. Hwang, & M. G. Frank (Eds.), APA handbook of nonverbal communication. Washington, DC: American Psychological Association.
4. Carlson, S. R., Katsanis, J., Iacono, W. G., & McGue, M. (1997). Emotional modulation of the startle reflex in twins: preliminary findings. Biological Psychology, 46(3), 235-246.
5. Cohn, J. F., & De la Torre, F. (2014). Automated face analysis for affective. The Oxford handbook of affective computing, 131.
6. Cohn, J. F., & Ekman, P. (2005). Measuring facial action by manual coding, facial EMG, and automatic facial image analysis. In J. A. Harrigan, R. Rosenthal & K. R. Scherer (Eds.), Handbook of nonverbal behavior research methods in the affective sciences (pp. 9-64). New York: Oxford.
7. Cordaro, D. T. (2014). Universals and Cultural Variations in Emotional Expression. University of California, Berkeley.
8. Cordaro, D. T., Keltner, D., Tshering, S., Wangchuk, D., & Flynn, L. M. (2016). The voice conveys emotion in ten globalized cultures and one remote village in Bhutan. Emotion, 16(1), 117.
9. Derbaix, C. M. (1995). The impact of affective reactions on attitudes toward the advertisement and the brand: A step toward ecological validity. Journal of Marketing Research, 32, 470-479.
10. Ekman, P. (1973). Cross-cultural studies of facial expression. Darwin and facial expression: A century of research in review, 169222, 1.
11. Ekman, P. (1980). The face of man. New York, NY: Garland Publishing, Inc.
12. Ekman, P. (1984). Expression and the nature of emotion. Approaches to emotion, 3, 319-344.
13. Ekman, P. (1992a). An Argument for Basic Emotions. Cognition and Emotion,

6(3/4), 169-200.

14. Ekman, P. (1992b). Are There Basic Emotions? Psychological Review, 99(3), 550-553.
15. Ekman, P. (1994a). All Emotions Are Basic. In Ekman, P. & Davidson, R. (Eds.), The Nature of Emotion: Fundamental Questions (pp. 15-19). New York: Oxford University Press.
16. Ekman, P. (1994b). Moods Emotions and Traits. From P. Ekman, & RJ Davidson (Eds.). The Nature of Emotion: Fundamental Questions (pp. 56-58).
17. Ekman, P. (1997). Emotion families. In Rauch, I. & Carr, G. F. (Eds.), Semiotics Around the World: Synthesis In Diversity (pp. 191-193). Berlin: Mouton de Gruyter.
18. Ekman, P. (1999). Basic Emotions In T. Dalgleish and T. Power (Eds.) The Handbook of Cognition and Emotion Pp. 45-60.
19. Ekman, P., & Cordaro, D. (2011). What is meant by calling emotions basic. Emotion Review, 3(4), 364-370.
20. Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. Journal of personality and social psychology, 17(2), 124.
21. Ekman, P., & Friesen, W. V. (1982). EMFACS Unpublished manuscript, San Francisco, CA.
22. Ekman, P., & Friesen, W. V. (1986). A new pan-cultural facial expression of emotion. Motivation and emotion, 10(2), 159-168.
23. Ekman, P., & Heider, K. G. (1988). The universality of a contempt expression: A replication. Motivation and emotion, 12(3), 303-308.
24. Ekman, P., Friesen, W. V., & Ellsworth, P. (2013). Emotion in the human face: Guidelines for research and an integration of findings. Elsevier.
25. Ekman, P., Friesen, W.V., Hager J.C. (2002a). Facial Action Coding System. 2nd ed., Salt Lake City: Research Nexus eBook.
26. Ekman, P., Friesen, W.V., Hager J.C. (2002b). Facial Action Coding System: Investigator's Guide. 2nd ed., Salt Lake City: Research Nexus eBook.
27. Ekman, P., O'Sullivan, M., & Matsumoto, D. (1991a). Confusions about context in the judgment of facial expression: A reply to "The Contempt Expression and the Relativity Thesis." Motivation and Emotion, 15, 169-176.
28. Ekman, P., O'Sullivan, M., & Matsumoto, D. (1991b). Contradictions in the study of contempt: What's it all about? Reply to Russell. Motivation and Emotion, 15(4), 293-296.

29. Ekman, P., Sorenson, E. R., & Friesen, W. V. (1969). Pan-cultural elements in facial displays of emotion. Science, 164(3875), 86-88.
30. Elfenbein, H. A. (2015). In-group advantage and other-group bias in facial emotion recognition. In Understanding facial expressions in communication (pp. 57-71). Springer, New Delhi.
31. Elfenbein, H. A., & Ambady, N. (2002a). Is there an ingroup advantage in emotion recognition? Psychological Bulletin, 128(2), 243–249.
32. Elfenbein, H. A., & Ambady, N. (2002b). On the universality and cultural specificity of emotion recognition: A meta-analysis. Psychological Bulletin, 128(2), 205–235.
33. Elfenbein, H. A., & Ambady, N. (2003a). Cultural similarity's consequences: A distance perspective on cross-cultural differences in emotion recognition. Journal of Cross-Cultural Psychology, 34(1), 92–110.
34. Elfenbein, H. A., & Ambady, N. (2003b). When familiarity breeds accuracy: Cultural exposure and facial emotion recognition. Journal of Personality and Social Psychology, 85(2), 276–290.
35. Elfenbein, H. A., Beaupre´, M. G., Levesque, M., & Hess, U. (2007). Toward a dialect theory: Cultural differences in the expression and recognition of posed facial expressions. Emotion, 7(1), 131–146.
36. Frank, M. G., & Stennett, J. (2001). The forced-choice paradigm and the perception of facial expressions of emotion. Journal of personality and social psychology, 80(1), 75.
37. Hertenstein, M. J., Holmes, R., McCullough, M., & Keltner, D. (2009). The communication of emotion via touch. Emotion, 9(4), 566.
38. Hertenstein, M. J., Keltner, D., App, B., Bulleit, B. A., & Jaskolka, A. R. (2006). Touch communicates distinct emotions. Emotion, 6(3), 528.
39. Izard, C. E. (1971). The Face of Emotion (Appleton-Century-Crofts, New York). Google Scholar.
40. Izard, C. E. (1979). The maximally discriminative facial coding system. Newark: University of Delaware.
41. Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code?. Psychological bulletin, 129(5), 770.
42. Keltner, D., Tracy, J., Sauter, D. A., Cordaro, D. C., & McNeil, G. (2016). Expression of emotion. Handbook of emotions, 467-482.

43. Kendler, K. S., Halberstadt, L. J., Butera, F., Myers, J., Bouchard, T., & Ekman, P. (2008). The similarity of facial expressions in response to emotion-inducing films in reared-apart twins. Psychological medicine, 38(10), 1475-1483.
44. Kendler, K. S., Halberstadt, L. J., Butera, F., Myers, J., Bouchard, T., & Ekman, P. (2008). The similiarity of facial expressions in response to emotion-inducing films in reared-apart twins. Psychological medicine, 38(10), 1475-1483.
45. Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H., Hawk, S. T., & Van Knippenberg, A. D. (2010). Presentation and validation of the Radboud Faces Database. Cognition and emotion, 24(8), 1377-1388.
46. Levenson, R. W. (2003). Blood, sweat, and fears. Annals of the New York Academy of Sciences, 1000(1), 348-366.
47. Lewinski, P., den Uyl, T. M., & Butler, C. (2014). Automated facial coding: Validation of basic emotions and FACS AUs in FaceReader. Journal of Neuroscience, Psychology, and Economics, 7(4), 227.
48. Lewinski, P., den Uyl, T. M., & Butler, C. (2014). Automated facial coding: Validation of basic emotions and FACS AUs in FaceReader. Journal of Neuroscience, Psychology, and Economics, 7(4), 227.
49. Matsumoto, D. (1988). Japanese and Caucasian facial expressions of emotion (JACFEE) and neutral faces (JACNeuF). Intercultual and Emotion Research Laboratory, Department of Psychology.
50. Matsumoto, D. (2002). Methodological requirements to test a possible in-group advantage in judging emotions across cultures: Comments on Elfenbein and Ambady and evidence. Psychological Bulletin, 128, 236–242.
51. Matsumoto, D. (2007). Apples and Oranges: Methodological Requirements for Testing a Possible Ingroup Advantage in Emotion Judgments from Facial Expressions.
52. Matsumoto, D., & Assar, M. (1992). The effects of language on judgments of universal facial expressions of emotion. Journal of Nonverbal Behavior, 16(2), 85-99.
53. Matsumoto, D., Hwang, H. C., Harrington, N., Olsen, R., & King, M. (2011). Facial behaviors and emotional reactions in consumer research. Acta de Investigacion Psicologica (Psychological Research Records), 1(3), 441–453.
54. Matsumoto, D., Keltner, D., Shiota, M. N., O'Sullivan, M., & Frank, M. (2008). Facial expressions of emotion. Handbook of emotions, 3, 211-234.
55. Matsumoto, D., Olide, A., & Willingham, B. (2009). Is there an ingroup

advantage in recognizing spontaneously expressed emotions?. Journal of Nonverbal Behavior, 33(3), 181.

56. Matsumoto, D., Olide, A., Schug, J., Willingham, B., & Callan, M. (2009). Cross-cultural judgments of spontaneous facial expressions of emotion. Journal of Nonverbal Behavior, 33(4), 213.
57. Matsumoto, D., Yoo, S. H., & Fontaine, J. (2008). Mapping expressive differences around the world: The relationship between emotional display rules and individualism versus collectivism. Journal of cross-cultural psychology, 39(1), 55-74.
58. Olszanowski, M., Pochwatko, G., Kuklinski, K., Scibor-Rylski, M., Lewinski, P., & Ohme, R. K. (2015). Warsaw set of emotional facial expression pictures: a validation study of facial display photographs. Frontiers in psychology, 5, 1516.
59. Papadatos, C., Alexiou, D., Nicolopoulos, D., Mikropoulos, H., & Hadzigeorgiou, E. (1974). Congenital hypoplasia of depressor anguli oris muscle: A genetically determined condition?. Archives of Disease in Childhood, 49(12), 927-931.
60. Peleg, G., Katzir, G., Peleg, O., Kamara, M., Brodsky, L., Hel-Or, H., Keren, D. & Nevo, E. (2006). Hereditary family signature of facial expression. Proceedings of the National Academy of Sciences, 103(43), 15921-15926.
61. Romashov, V. & Shakhraichuk, I. (2017). Stimuli quantity: More practice – higher agreement? Unpublished report.
62. Romashov, V. (2016). Detecting Aggressive Emotional States. Unpublished Master's thesis. Taras Shevchenko National University of Kyiv, Kyiv.
63. Romashov, V. (2018). Fixed Choice format: response option matters? Unpublished report.
64. Romashov, V. (2018). Within and between-subject design in emotion judgment studies. Unpublished report.
65. Romashov, V., Shakhraichuk, I., & Daniluk, I. (2016). Response Forman Comparison: Fixed Choice vs. Forced Choice, I. Unpublished report.
66. Russell, J. A. (1980). A circumplex model of affect. Journal of personality and social psychology, 39(6), 1161.
67. Russell, J. A. (1991a). Negative results on a reported facial expression of contempt. Motivation and Emotion, 15(4), 281-291.
68. Russell, J. A. (1991b). Rejoinder to Ekman, O'Sullivan, and Matsumoto. Motivation and emotion, 15(2), 177-184.
69. Russell, J. A. (1991c). The contempt expression and the relativity thesis.

Motivation and emotion, 15(2), 149-168.

70. Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. Psychological bulletin, 115(1), 102.
71. Russell, J. A., & Bullock, M. (1986). Fuzzy concepts and the perception of emotion in facial expressions. Social Cognition, 4(3), 309-341.
72. Sauter, D. A., Eisner, F., Ekman, P., & Scott, S. K. (2010). Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. Proceedings of the National Academy of Sciences, 107(6), 2408-2412.
73. Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. Speech communication, 40(1-2), 227-256.
74. Sorenson, E. R. (1975). Culture and the expression of emotion. In T. R. Williams (Eds.), Psychological anthropology (pp. 361-372). Chicago: Aldine.
75. Sorenson, E. R. (1976). The edge of the forest: Land, childhood and change in a New Guinea protoagricultural society. Washington, DC: Smithsonian Institution Press.
76. Van Der Schalk, J., Hawk, S. T., Fischer, A. H., & Doosje, B. (2011). Moving faces, looking places: validation of the Amsterdam Dynamic Facial Expression Set (ADFES). Emotion, 11(4), 907.
77. Wagner, H. L. (1993). On measuring performance in category judgment studies of nonverbal behavior. Journal of Nonverbal Behavior, 17(1), 3-28.
78. Wagner, H. L. (1997). Methods for the study of facial behavior. The psychology of facial expression, 31-54.
79. Wagner, H. L. (2000). The accessibility of the term “contempt” and the meaning of the unilateral lip curl. Cognition & Emotion, 14(5), 689-710.
80. Wallbott, H. G. (1998). Bodily expression of emotion. European journal of social psychology, 28(6), 879-896.
81. Warren, G., Schertler, E., & Bull, P. (2009). Detecting deception from emotional and unemotional cues. Journal of Nonverbal Behavior, 33, 59–69.
82. Wingenbach, T. S., Ashwin, C., & Brosnan, M. (2016). Validation of the Amsterdam Dynamic Facial Expression Set–Bath Intensity Variations (ADFES-BIV): A set of videos expressing low, intermediate, and high intensity emotions. PloS one, 11(1), e0147112.
83. Frank, M. G., & Ekman, P. (1993). Not all smiles are created equal: the differences between enjoyment and nonenjoyment smiles. Humor: International Journal of

Humor Research.

84. Zeinstra, G. G., Koelen, M. A., Kok, F. J., & de Graaf, C. (2009). Facial exressions in school-aged children are a good indicator of 'dislikes', but not of 'likes'. Food Quality and Preference, 20, 620-624.

Internet sources

http://www.qualtrics.com

http://www.vassarstats.net/rdiff.html

Lists of tables, diagrams and graphs

Annexes

Annex 1: Mean H_u and P_c results in Experiment

		P_c	H_u
Anger	Forced	0.012	0.648
	Fixed	0.008	0.543
Contempt	Forced	0.017	0.601
	Fixed	0.010	0.438
Disgust	Forced	0.018	0.703
	Fixed	0.016	0.662
Fear	Forced	0.012	0.558
	Fixed	0.010	0.572
Happiness	Forced	0.015	0.897
	Fixed	0.014	0.831
Neutral	Forced	0.017	0.824
	Fixed	0.014	0.808
Sadness	Forced	0.017	0.780
	Fixed	0.015	0.690
Surprise	Forced	0.018	0.756
	Fixed	0.018	0.748

Source: own elaboration

Warsaw, ……

Name and surname: ……..

Student's ID no.: …….

DECLARATION

I declare that the Master's Dissertation titled.:

……
……
……
……

which was submitted at Kozminski University, was written by me alone and has not previously been the subject of any official procedure associated with applying for a higher education diploma which confirms obtaining a professional title.

I also declare that the present dissertation does not infringe copyrights under Act of 4 February 1994 on Copyright and Related Rights (Journal of Laws of 2016, item 666) or personal rights protected by law.

Concurrently, I am aware that the content of the dissertation will be verified with the antiplagiarism program which cooperates with the national repository of written dissertations.

/Student's signature/

Warsaw, ……..

Name and surname of the Supervisor: ……

DECLARATION

I declare that the Master's Dissertation titled:
……
……
……
……

authored by the student …… (student's ID no. ……..) and submitted at Kozminski University, was prepared under my supervision.

Concurrently, I acknowledge that the content of the Master's Dissertation will be verified with the anti-plagiarism program which cooperates with the national repository of written dissertations.

/Supervisor's signature/