UNIVERSITY OF PADOVA

Department of General Psychology

Master Degree in Cognitive Neuroscience and Clinical Neuropsychology

Final dissertation

Detection of simple and complex deceits through facial micro-expressions: a comparison between human beings' performances and machine learning techniques

Supervisor
Professor Cristina Scarpazza

Co-supervisor
Professor Merylin Monaro
Professor Nicolò Navarin*

Candidate: **Lorenzo Silla**
Student ID number: **2002397**

Academic Year 2021/2022

*Department of Mathematics "Tullio Levi-Civita"

TABLE OF CONTENTS

ABSTRACT

Micro-expressions have gained increasing interest in the last few years, both in scientific and professional contexts. Theoretically, their emergence suggests ongoing concealments, making them arguably one of the most reliable cues for lie detection (e.g., Yan, Wang, Liu, Wu & Fu, 2014; Venkatesh, Ramachandra & Bours, 2019). Given their fast onset, they result almost imperceptible to the eye of an untrained subject, making it necessary to work on automatic detection tools. Machine learning models have shown promisingly results within this domain; thus, the aim of the study at hand, was to compare the performances human judges and machine learning models obtain on the same dataset of stimuli. Regrettably, machine learning performances have ended up being around the chance level, positing the question of why previous and a-like studies have collected better results. Briefly, insights on how to properly organize an experimental paradigm and collect a dataset for lie detection studies are discussed, while concluding that among other several necessary cues it is still crucial to consider micro-expressions when dealing with lie detection procedures.

1. INTRODUCTION

He who has eyes to see and ears to hear may convince himself that no mortal can keep a secret. If his lips are silent, he chatters with his fingertips; betrayal oozes out of him at every pore (Freud, 1959, p. 94).

This quote from the "Collected Papers" of Sigmund Freud (1959) highlights the human being's belief for easily catching others lying. As we will discuss later on, cues are actually present in our body resembling our deceptions. Nevertheless, differently from what Freud may have suggested, human beings are not good lie detectors. Several studies attempted at assessing the proportion of correct guesses when an individual is asked to spot someone lying, and on average a human judge does not diverge significantly from the chance level indeed (e.g., Jordan et al., 2019). In sum, even if a variety of cues to deception are there, surprisingly they can be seen by some but not by most. Why this deficiency is so evident?

The first line of research that tried to fully meet this question was conducted by Ekman in 1990. He gave five explanations for understanding why we fail in spotting liars. The first one takes an evolutionary point of view. It is assumed that our ancestors would not particularly benefit from deceiving their fellows: in a community where survival relied upon cooperation being caught lying might be deadly. It is no surprise that in a stable, small and close group any attempt at deceptive communication might easily lead to social problems, translating into an augmented likelihood of just casual subtle and low-frequency lies (Cheney & Seyfarth, 1990). In addition, privacy needs were poorly expressed in these societies, meaning that for an individual was nearly impossible to not share or hide his thoughts and experiences. Briefly, being talented in detecting lies was not adaptive in any sense in such circumstances.

Even if we admit that our ancestors were not good (and interested too) in lying, the question of why we do not evolve these skills while growing up remain unsolved. Ekman (1990) assumed that our parents choose to teach us to not catch their lies. The reasoning behind that might be the privacy that parents require when raising a child: parents may be concerned about concealing their behaviours and details about their everyday plans. In this regard, the most evident (but not exclusive) aspects asking for concealment are those correlated to sexual activities.

Another explanation for our poor deception detection skills is the costs following the unfolding of a lie: being trustful allows us to both protect social status from potential harmful consequences and preserve cognitive energy. Nevertheless, being always suspicious might undermine the likelihood of establishing intimacy with other human beings (Ekman, 1990). Moving on, Ekman posited another social issue that has to be considered to properly answer our leading question. It happens in certain circumstances that an individual might indeed prefer to be misled more than knowing the truth (Tooby & Cosmides, 1994; Ekman, 1990): this strategy allows to postpone or avoid confronting with an unpleasant truth, and it leads to unconsciously overlooking liar's mistakes. A very last consideration belonging to the social domain refers to the need of being polite throughout our interactions with peers: we do not want to steal information that is concealed by the sender. In fact, it is acceptable to acknowledge that in most cases we are just interested in storing the message the sender plan to communicate to us: we do not want to look for concealed information because doing so might impact the smoothness of the communication itself (Goffman, 1974; Ekman, 1990).

Having said that, all the discussed assumptions still do not explain why low deception detection scores are observed even in trained subjects, and members of the criminal justice. These individuals have no reason to accept being misled, they are conversely trained and instructed to look for information that is concealed. Nevertheless, they still get results that are close to the chance level (Ekman, 1990). However, two possible issues inside this reasoning are there: the high rate of liars the professionals are presented with, and even the inadequate feedback they receive. Having to deal with a high-base rate of liars does not allow trained individuals to become enough receptive to subtle behavioural cues. Moreover, it is noteworthy to stress that usually, these professional figures are interested in getting a piece of evidence to nail the liar more than understanding how to spot the liar per sé. In addition, these individuals usually receive only delayed feedback: once they make a mistake and have the chance to understand they have done it wrong, it is too late to correct the judgment.

Taking all these findings as a whole, it is clear how nowadays different professional contexts are in demand for strategies meant at ameliorating the detection of liars. For instance, as well as job recruiters might be interested in unfolding the truth in attitudinal tests to hire the best person for a job position, a police officer might be concerned with solving crimes by identifying the actual facts. Even in the psychiatric setting, understanding the goodness of a statement concerning a patient's well-being appears to be critical (Ford, 2005). Thankfully, since the traditional polygraph appear in the first half of the nineteenth century, a series of lie detection systems have been developed to assist us within professional contexts. Recently, some proposed tools meant at improving lie detection performances have incorporated artificial intelligence (AI) techniques (Oravec, 2022), as the automated analysis of facial micro-expressions that are the central focus of this work.

1.1. Do people look for effective cues to detect deception?

We already discussed that the typical rates of lie-truth discrimination are just slightly above 50% when conducted by human beings (e.g., Bond & DePaulo, 2006). Ekman (1990) handled the issue by taking an evolutionary, developmental, and eventually social perspective, stating the accountability of our parents and social values. However, given that cues to spot deception are available, it is possible to speculate on two more pragmatic explanations. Firstly, it might be the case that people are unable to catch lies because they rely on invalid cues. Alternatively, these hints might be too dearth and with a small associated effect size leading to poor accuracy (Steringlanz et al., 2019). To deal with this question Hartwig and Bond (2011) measured within a meta-analysis the overall correlation between deception's perceived cues and actual cues. After examining 66 cues across 153 samples they obtained a correlation of $r = .59$, namely a moderate to a strong association. Taking into account only those studies in which perceived and actual signs were measured on the same group of perceivers and senders the correlation raises to $r = .72$ (i.e., strong relationship). Researchers conclude that besides low levels of accuracy people usually look for the right cues: that is, incorrect hints are not considered and limited detection accuracy can be attributed mostly to the fact that valid cues to deception are not always reliable (Hartwig & Bond, 2011).

One last point to be noted is that cues on which actually people rely when attempting at catching lies are different from those they claim to rely on. In the same meta-analysis, Hartwig and Bond found that receivers are unaware of hints they pay attention to: when people are asked to think about cues they mostly rely on, avoiding eye gaze is frequently posited as a first and central hint; interestingly lack of eye contact is only weakly correlated ($r = .15$) to perceivers' judgment of deceptiveness (Hartwig & Bond, 2011). This is consistent with general findings from psychological literature positing that people are often misguided when making claims about their internal cognitive processes (Nisbett & Wilson, 1977). In sum, it seems that people are not aware of the cues they use to spot liars.

It is now evident that social, developmental, and evolutionary explanations were only one part of the entire picture. Alongside our innate sense of avoidance towards admitting someone is deceiving us, researchers have highlighted the low external validity showed by deception's cues: reliable cues indeed exist, although their low generalizability to different senders' mental states and to contextual changes do not allow to always obtain sure hints, at least when the detector is a human being.

1.2. What defines deception and the specific case of lying: why do we lie?

Lies are ubiquitous in today's human communication. Ranging from white to more malicious lies, people are used to being deceptive when interacting with others. The specific biological evolution of our brain has allowed us to become fine liars by differentiating our species in the use of deceptive tactics with social manipulation intent (Abe, 2009). The size of the human neocortex is certainly and crucially involved in the availability of such competencies: in every modern primate's species, it has been observed a positive correlation between the size of the neocortex and the use of deceptive communication (Byrne & Corp, 2004; Figure 1.1).

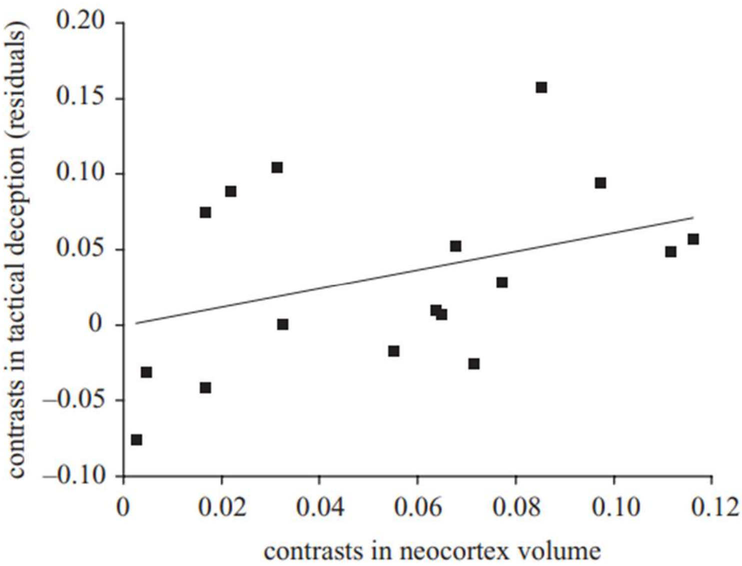


Figure 1.1. *Correlation between tactical deception's frequency ("acts from the normal repertoire of the agent, deployed such that another individual is likely to misinterpret what the act signifies, to the advantage of the agent") and volume of the neocortex in different primates' species obtained through stepwise multiple regression analyses. The x-axis shows an index of neocortical size, and the y-axis shows an index of deception usage. Independent contrasts were used to avoid a taxonomic bias (Byrne & Corp, 2004).*

According to Abe's definition (2009, 2011), deception can be defined as a psychological process by which an individual consciously and deliberately tries to convince another person to accept as true what the liar knows to be false. Notably, within this definition it is implied that unconscious and mistaken acts of misleading cannot be considered a form of deception: for most researchers, a core feature of deception is that it is an intentional and planned act. This statement helps us in stressing the important distinction between lying and so-called false beliefs. For instance, a boy that is asked to recollect a memory trace concerning a car's appearance may misremember it, even if he is sure that is reporting a true statement. In such a case, the boy's report is not claimed to be a lie (Ceci & Bruck, 1998).

Another implication of such a definition is that if two people contradict each other, this does not directly mean that one of them is lying. How this feature affects criminal cases is clear: in a setting in which two people are reporting two contradictory statements there is no room to be assured that at least one of them is lying. It could be the case that none of the witnesses is lying but one is misremembering the event (Vrij, 2009).

The emerging central role of intentionality underlies one more feature of deception definition: apart from being an intentional act, deception must be always understood from the perspective of the deceiver, and not from the factuality of the report. A lie is considered as such when the sender believes that what he or she is saying is false. Thus, an actual truth remarkably could be a lie. Moreover, considering the deceiver's perspective entails that a statement that initially was considered a lie might lose its status over time (Pickel, 2004). In a study by Pickel (2004), a group of participants were instructed to report true or false statements about a movie they had been presented with. A week later they were then asked to recall some details of what they had watched. Those who a week before were asked to report false statements gave more wrong details than the subjects asked to produce a true report. Just the brief report of false statements had been enough to change the memory traces that these individuals have ended up perceiving as true.

Lastly, not all researchers are in accordance to include in the deception domain one more specific and unusual subcategory: self-deception. It is not uncommon that people delude themselves in a way that they become convinced (unconsciously and not rationally) about the goodness of a wrong reasoning. People use indeed self-deception both as a protective or as an avoiding strategy. Thus, self-deception may have both positive and negative consequences (Lewis, 1993). On the one hand, people may ignore potentially life-threatening facts (e.g., body symptoms) and convince themselves that an intervention is not needed. On the other hand, self-deception may be used to protect self-esteem. For instance, this strategy would be useful to avoid dealing with the affection following being rejected by a potential partner (Vrij, 2009).

1.3. Cognitive process and brain areas involved in the act of lying

After having dealt with deception's definition, it appears clear how the pure act of deceiving has to involve multiple cognitive processes. Different theories are aimed at understanding how behavioural and verbal expressions are controlled when telling a lie. Even if there are different explanations about how mental processes interact with each other when producing a lie, theorists agree on giving a crucial role to executive functions' involvement (Gombos, 2006). Executive functions account for the active management and control mechanism of thought, thus they are involved in cognitive activities such as directed attention and metacognition, working memory and cognitive inhibition (Fernanedz-Duque, Baird & Posner, 2000, Gombos, 2006). In usual situations, liars are required to deal with two simultaneous processes: they must build a new trace of information (i.e., the lie) while also withholding a factual and accurate trace (i.e., the truth). It is assumed that the liar knows which is the actual truth, which leads to a form of baseline or predominant response. Crucially, the correct information is expected to be communicated not only by an honest subject that is presented with the very same question but even by a liar that is distracted or tired. Thus, lying required additional cognitive processing that engages the executive and prefrontal system more than a truth-telling does (Zeki et al., 2004). As a confirmation of what has been just said, Zuckerman and colleagues (1981) argued that lying requires more cognitive effort than truth-telling because liars have to maintain simultaneously internal consistency (to be reassured that the made-up features of the fabricated story are coherent) and external consistency (to be reassured that the fabricated features of the fictional product align with other's understanding). Within the theoretical framework of a cognitive complexity theory, Zuckerman assumed that the cognitive workload due to the deception production should be usually accompanied by some behavioural and bodily changes, such as pupil dilations, longer response latencies and more speech hesitations (Zuckerman et al., 1981).

Although observing an augmented cognitive load when telling a lie is the norm, some exceptions are present: on some occasion, it seems that lying is a no more difficult task than telling the truth. For instance, when producing a true statement entails extended explanations or detrimental consequences, a lie may be smoother to express than a truth (Vrij, 2000). There are confirmations to this assumption coming from the same meta-analysis by DePaulo and colleagues (2003): they observed specific factors that may increase or decrease the effort expected to be put in when producing a lie. The opportunity to plan the deceptive statements and the duration of it have been indicated as being moderating factors of the cognitive load correlated to deception's production. In particular, they reported that having less time to arrange a lie is related to a wider amount of cognitive effort and that nevertheless the longer the duration of the deception is the more difficult its output will be (DePaulo et al., 2003).

Developmental psychology studies' conclusions are yet another evidence of the strong involvement of executive functions in deception. Firstly, some theorists have pointed to the acquisition of "the theory of mind" (Wellman, 1992) as needed to develop the ability to deceive (e.g., Wimmer & Parner, 1983). Theory of mind represents the ability of one individual to acknowledge the possibility that one another has different beliefs, desires, and intentions from your own one. Studies on deception skills have evidenced that 3-year-old children lack the ability to deceive compared with 4- or 5-year-old ones (e.g., Peskin, 1992). This is not surprising due to the perspective concerning the acquisition of theory of mind by age of 4. However, the view that the strengthening of deceptive ability is only owed to the development of a concept (e.g., theory of mind) rather than other functional changes has been questioned. In fact, the positive correlation between high scores on the theory of mind task and those requiring executive functions appears to be the actual solution: the observed correlation between theory mind "false beliefs tasks" and card sorting (that relies heavily on the goodness of executive functions) stresses that executive functions may be a prerequisite for the formation of false beliefs (Perner, Lang & Kloo, 2002). In sum, in children (as in adults), deceptive ability's maturation is characterized by a growth that parallels the development of executive functioning and especially of cognitive inhibition, rather than being just associated with the acquisition of a concept understanding (Gombos, 2003).

1.3.1. Cognitive theories of deception

The first theory that attempted at giving a cohesive understanding of the processes underpinning deception was proposed by Zuckerman and colleagues (1981). Within the Four-Factor theory, they stressed some characteristics that differentiate liars from truth-tellers. Firstly, they were assured that generalized arousal is greater when lying, and that emotions observed when lying are due to guilt. About cognitive processes, Zukerman's conclusions were yet to be exhaustive: they conclude that cognitive aspects correlated to deception are much more complex than those expected during truth-telling and that extensive control of verbal and nonverbal behaviours is needed to avoid getting caught when telling a lie (Gombos, 2003).

Later on, the Interpersonal Deception Theory (IDT; Buller & Burgoon, 1996) explained deception as a matter of two-way interactive communication. Under this theoretical framework, the liar is seen as producing a deceptive statement while monitoring the target individual for a sign of suspiciousness. In such an overseeing, the liar uses the signs coming from the target as tools to adapt his or her behaviour while the lie is passing off. Conversely, the target will attempt to catch verbal and non-verbal cues with the scope of spotting potential incoming deceptive messages. Spotting a lie may result facilitated when the deceiver experiences a cognitive overload: his or her inhibition processes (responsible for the concealment of the truth) may become ineffective leading to the outflow of more evident cues suggesting deception. Most targets otherwise would not usually spot deception. Briefly, IDT theory suggests that deception is a mentally taxing activity in which the involvement of executive cognitive resources is crucial: the liar exploits a metacognitive regulation and executive attention (Fernandez-Duque et al., 2000), alongside behavioural and cognitive inhibition (i.e., suppressing nonverbal and verbal cues; Buller & Burgoon, 1996).

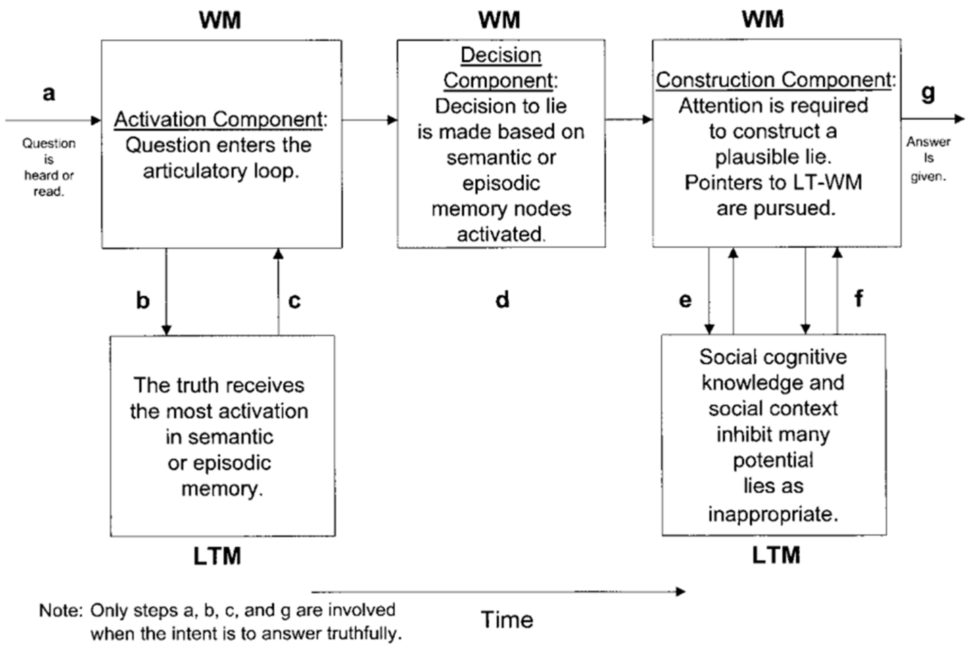


Figure 1.2. *The progressive processes of the activation-decisions-construction model are depicted. Its trigger is an input question that is asked or read. As a final step, after recollecting all the needed information and context-related knowledge, the lie output is produced (Walczyk et al., 2000).*

The second theoretical framework of interest was developed by Walczyk and colleagues (2003) who tested a cognitive-based theory of deception: the activation-decision-construction model (ADCM) of lying. Three cognitive events are taken into account herein, usually occurring in the following order: an activation component, a decision component and a construction component (Figure 1.2). The first trigger within this model is the activation of the working memory component that allows retrieving the knowledge of the actual truth. The very next event is the decision-making process concerning whether (or not) the subject should lie, and how to possibly produce the deceit itself. Notably, within the unfolding of these stages, inhibition is required to suppress the outgoing of details regarding the truth. As a final process, deception construction requires attention to help retrieve information about the social context and those useful key elements stored in memory to form plausible lies (Walczyk, 2000).

The last model considered was proposed by Mohamed and colleagues (2006). They presented a model of deception that aims at integrating the cognitive and neural correlates related to lying. The deceptive response is triggered by hearing or reading a question, others six stages related to mostly distinct brain areas' activation (that may occur simultaneously) follow. These subsequent steps include the comprehension of the question, the recall of the events concerning the same question and the judgment and establishment of response (at this stage inhibition of preponderant output is included). Throughout the unfolding of these steps, the sender may experience senses of fear, anxiety, apprehension, guilt, or joy which are an integral part of the final verbal response production. Not surprisingly, as Mohamed and colleagues noted, the only stage that may differentiate liars from truth-tellers is the one concerning the response planning and, most importantly inhibition: both truth-tellers and liars need to understand and store the languages input, retrieve memory traces of an event, and produce a verbal response. Thus, albeit the verbal contents of lies tend to differentiate from those of truth statements, Mohamed and colleagues assumed that the same brain areas are involved in carrying out both the two and opposite types of statements. They indicated Broca's area and the precentral gyrus as activating during the verbal response, both in truth-tellers and liars (Mohamed et al., 2006).

By examining the most important model concerning deception cognition the prominence of executive processing has been unfolded. Researchers from the lie detection domain and developmental and neurocognitive studies emphasize those core features of executive functions that are crucial for deception, such as inhibition, attention, metacognition, and guidance by the working memory. The proofs concerning this centrality are even consistent with the above: by assuming that the access to executive processes (i.e., cognitive resources) is limited, it is reasonable to accept that lying become increasingly difficult as time goes by and as the deceiver becomes more and more cognitively overloaded (Gombos, 2006).

1.3.2. Insights into the neuroscience of deception

Functional neuroimaging studies cannot but be useful means to study the brain correlates of deception. Nonetheless, differently from other types of cognitive processes, the experimental settings meant for studying deception might finish having an extremely low ecological validity. Simulated deception in laboratory settings cannot be viewed indeed as having the same cognitive- and emotional-load as deception has in real life. However, experimental tasks that attempt at reproducing as much as possible the actual processes involved in deception do exist (Abe, 2011).

As corroborated by the previously discussed cognitive models, it seems that the frontal executive system is associated with deception (e.g., Abe, 2009, Christ et al., 2009). Areas that have been assumed to play a key role in deception production are the dorsolateral prefrontal cortex, the ventrolateral prefrontal cortex, the anterior prefrontal cortex and the anterior cingulate cortex. These regions have indeed been implicated in different cognitive processes covering deception; for instance, the dorsolateral prefrontal cortex is implicated in working memory-related monitoring, response selection and general cognitive control (Rowe et al., 2000; MacDonald et al., 2000).

The cognitive processes underlying deception are really broad, making it challenging to study the actual functions of interest: although fMRI is an essential tool to measure the brain activity of deception, its conclusion may lack specificity. In that respect, a robust solution is the loss-of-function studies: the observation of the functionality of brain-damaged neuropsychological patients is an effective aid to avoid making overly general conclusions. Abe and others (2009) resolved to examine the ability to tell a lie in Parkinson's (PD) patients; interestingly, some previous works on PD noticed that this kind of patient shows surprisingly low levels of deceptive behaviours (Ishihara & Brayne, 2006; Menza, 2000). Abe and colleagues attempted at evaluating this deficit under a different optic: it could be the case that PD patients end up being unable to deceive due to pathological changes in their brain, rather than being reluctant to deceive other people. This hypothesis proved to be correct eventually. PD patients have indeed difficulty in making deceptive responses within cognitive tasks if compared with the standard scores of control subjects. Using resting-state 18F-fluorodeoxyglucose (FDG) PET the deceptive behaviours' avoidance was significantly associated with decreased metabolic rates in the left dorsolateral and right anterior prefrontal gyrus, suggesting the plausible strong involvement of these areas within deceptive communication (Figure 1.3, Abe, 2011).

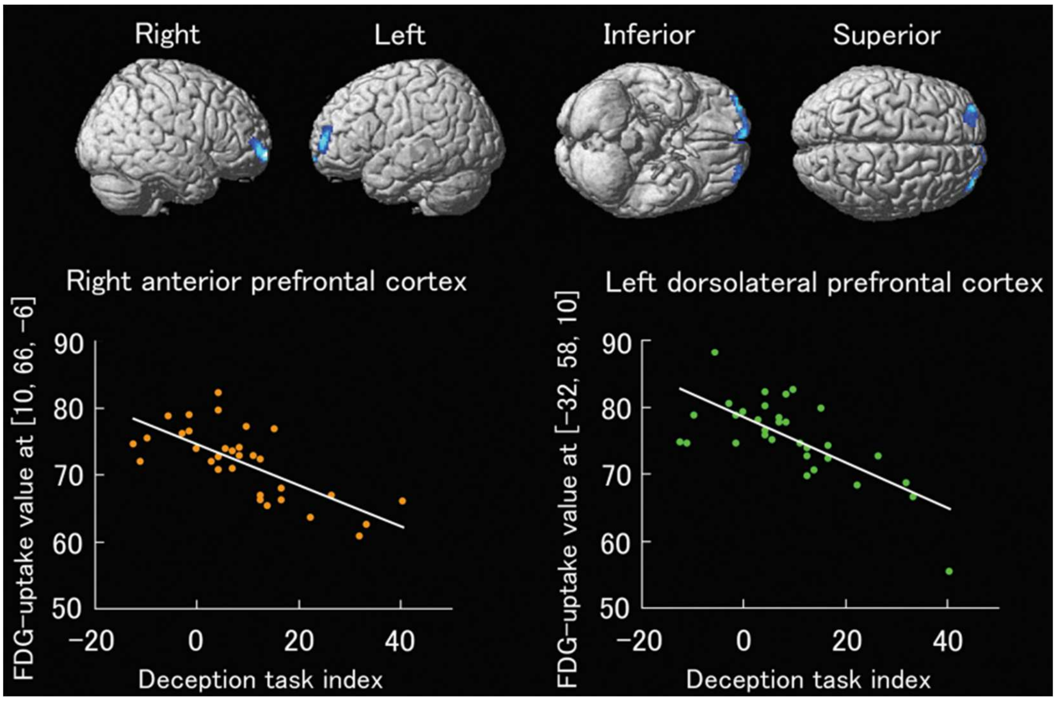


Figure 1.3. *Firstly, brain regions showing a significant correlation between scores in a deceitful task and regional cerebral glucose metabolism in PD disease are reported. Beneath, the scatterplot shows the association between performance in a deceitful task and regional resting brain glucose metabolism in PD patients. They show the negative correlation between deception task indices and the fluorodeoxyglucose (FDG) -uptake values in both the right anterior prefrontal cortex and left dorsolateral prefrontal cortex. In sum, it was highlighted a significant negative correlation between the deception task index and the metabolic rates of the two reported brain regions. Finally, it is noteworthy to stress that results are masked by the contrast of control subjects against PD patients. Thus, these two areas were found to show hypo-metabolism in patients if compared to control subjects (Abe, 2009).*

Taking into consideration the low ecological validity of experimental-based deception, Greene and Paxton elaborated a unique experimental paradigm for increasing adherence to real life: they proposed an fMRI-based experiment in which subjects could gain a monetary income by accurately predicting the outcomes of computerized coin flips (2009). Two were the experimental conditions. In some trials, participants submitted their predictions in advance. In others, subjects were instead rewarded based on self-reported predictions made after the flips. This second group of subjects could get monetary earnings by fraudulently reporting the accuracy of their predictions. An increased prefrontal activity was observed in the delayed-report condition, both in those subjects that resolved to be dishonest and in those that refrain from the temptation of adapting their predictions to factual knowledge. In the control condition in which the possibility of acting dishonestly was null, the (honest) subject did not exhibit such control-related activity. Greene and Paxton agreed that the absence of prefrontal activity in honest behaviours is closely linked to an actual absence of temptation, rather than active resistance to it (Greene & Paxton, 2009).

A resulting issue is understanding which is the exact role of the prefrontal cortex (PFC) in deceptive behaviour, given that it is too simplistic to interpret PFC activity only in terms of working memory or response inhibition. As proposed by Greene and Paxton one credible explanation is that activity may underlie the processes of actively deciding whether to lie independently from the choice made. For what concern dishonest subjects, PFC activity may instead reflect the attempt to resist temptation (Greene & Paxton, 2009).

With the scope of integrating the findings by Greene and Paxton, Abe (2011) proposed to focus even on subcortical structures, such as the amygdala and the ventral striatum. Both these areas were found to be activated in neuroimaging studies of deception (Abe et al., 2007; Baumgartner et al., 2009). In these subcortical areas, reward-seeking and emotional regulation for deceptive behaviour may be processed, as well as those personality traits that may be assumed to determine honesty or dishonesty. In sum, the PFC might play a key role in top-down processes such as evaluating whether to tell a lie or not in complex social scenarios. Then, again, the PFC may feature in the output of deceptive behaviours which is more demanding than truth-telling in some circumstances. Triggering deceptive conduct may be instead achieved in subcortical areas through the onset of motivational and morality-related components (Abe, 2011).

1.4. Drawing distinctions between different types of lies

Alongside the definition and models conceptualizing lies, scholars have focused on understanding the different ways in which people categorize lies (Bryant, 2008). Generally speaking, it is accepted to broadly categorize lies in at least two broad categories, given that people agree on the existence of some lies that are less severe than others (Seiter et al., 2002; Turner et al., 1975): white lies and exploitive lies. A restricted group of lies is even considered acceptable (white lies) because they are trivial or, to the extreme, they spare people from being hurt by unnecessary truths. These benign lies can be viewed as a “social lubricant” which allows social interactions to evolve smoothly by avoiding disagreements and harming either person’s pride or self-image (Saxe, 1991). Exploitive lies, on the contrary, are harmful and unacceptable, they are indeed told with the scope of hurting someone and taking an advantage of them (Hopper & Bell, 1984).

Later on, Di Battista (1994) took another perspective in analysing this dichotomy: he stressed that the two categories split into lies told just to advantage your own self-interest (trust-violating) and lies told while bearing in mind other's feelings (tactful or white lies, Di Battista, 1994).

A slightly different, but yet fundamental, approach was followed by DePaulo and colleagues (1996) that resolved to divide lies' taxonomy into three branches: outright lies that are total falsehoods, exaggeration or truth overestimation, and subtle lies, which include the purposeful omission of details (DePaulo et al., 1996).

Outright lies (firstly defined by Ekman as falsification, 1997) are a unique type of deception in which the information conveyed is completely different from what the sender believes to be the truth. Outright lies can be typically encountered in forensic settings when guilty suspects consciously deny any involvement in a crime through the mean of misleading statements. Contrary to outright lies, exaggerations are lies in which facts are just over-or under-estimated. Thus, senders minimize or exaggerate their statements, rather than bending the truth per sé. A prototypical example of exaggeration is liars that embellish their well-being while presented with an interview (Vrij, 2008, DePaulo et al., 1996).

Remarkable features characterize the unique case of the subtle lies that can be viewed as literal truths designed to mislead. The smooth-running of subtle lies relies on different aspects, but usually, it entails hidden or narrower definitions concerning the matter of discussion or the concealing of information: in doing so, the senders may evade the initial question or omit relevant details. Not surprisingly, these characteristics make subtle lies' use the most common among the different types of deception, indeed concealing details is relatively easy besides being difficult to be detected. In addition, concealments are often viewed as being less negative than other forms of lying (Levine, 2001; Levine, Asada & Lindsey, 2003), thus liars should face fewer and softer repercussions from lie targets in case they are detected: digressing or reporting partially facts is easier to morally justify than reporting a comprehensive untruth (Vrij, 2008).

1.4.1. How do people categorize lies: a study

Even if lies' categorization exists, it is noteworthy to stress that this taxonomical debate is heavily dependent on an individual's perception, given that the evaluation of the goodness or viciousness of a lie relies on subjective lines of reasoning. This issue was dealt by Bryant while aiming to find coherence in lies' features (2008). He presented a series of college students with interviews and focus groups, with the scope of understanding how many categories of lies fell in participants' views (Bryant, 2008). Results indicated three main categories: white lies, real lies, and grey lies. Moreover, it was highlighted that people take into account five factors when analysing lies, namely the lie's intention (or motivations), its consequences, its benefits (i.e., whom a lie is intended to bring a benefit), the level of truthfulness and its amount of acceptability (i.e., the degree under which a lie was defined as permissible by the respondents; Figure 1.4).

Among the three categories, real lies were posited as the most serious and unacceptable form of lies, as they are fabricated with a conscious intention of deceiving others and they can cause receivers serious problems. Participants agreed on saying that real lies are told by completely fabricating the truth, leading this form of deception to have serious consequences and to be described as self-serving and malicious. White lies on the contrary were described as the least serious and nonetheless most acceptable: their intents are benign, and either altruistic or trivial at times. Participants agreed in stating that white lies are commonly used by most people and in many situations, given that they are a not harmful form of deception. Lastly, it was possible to acknowledge the existence of the third class of lies, namely grey lies: they were described as capturing the middle area between real and white lies while being a key solution for filling the gaps across people's ways of reasoning. Two groups resulted to be included in grey lies: ambiguous grey lies and justifiable grey lies. Ambiguous ones are lies that are open to multiple explanations, it is usual that people interpret them in different ways. Justifiable grey lies instead can be viewed as real lies which however can be considered justifiable in certain circumstances (Bryant. 2008).

Types of Lies	Factors: Intention	Consequences	Beneficiary	Truthfulness	Acceptability
Real Lies	Malicious Deliberate Deceptive Deceitful	Serious Direct	Self-Serving Egotistical	Complete Fabrication Blatant Untruth Zero Truth	Unacceptable Not Justified
White Lies	Benign Pure	Trivial Meaningless Harmless	Altruistic Other-Focused Protecting Helpful	Partial Truth Half Truth Bending the Truth Stretching the Truth	Acceptable Justified Expected Common
Gray Lies					
Ambiguous Gray Lies	Ambiguous Intention	Ambiguous Consequences	Ambiguous Beneficiary	Ambiguous Level of Truth	Open to Interpretation
Justifiable Gray Lies	Malicious	Direct	Self-Serving	Complete Fabrication	Justified Acceptable

Figure 1.4. *Within the table are reported the results based on participants' categorization in which real lies, white lies and grey lies were distinguished using as factors intention, consequences, beneficiary, truthfulness and acceptability. All the features of the three types of*

lies are taken singularly and described following the discussion of the same factors (Bryant, 2008).

1.4.2. The Cognitive-load associated with deception and its relation to lie's features

As it will be discussed more in-depth later on, one of the focuses of this work is catching potential differences between the detection scores associated with different types of lies. Two types of stimuli were used within the experimental paradigm of the current study, respectively complex images (i.e., images full of details, such as landscapes or urban settings) and simple images (i.e., objects or animals placed on a white background). Participants were just asked to describe these pictures while being left free to decide if producing a deceptive description or a factual one. So far, it is enough to stress that one of the assumptions of the current study was to observe the biggest cognitive load in presence of a complex figure's fictional description, differently from what it is expected to observe with a simple one. Lying indeed not always requires the same mental effort, and several aspects contribute to increasing and influencing this mental load (McCornack, 1997).

First, it has been already mentioned that just formulating the lie itself may result in being cognitively demanding, given that senders have to monitor their fabrications and inhibit the reality of facts. Thus, handling all the details of a complex figure might require a wider effort when producing a false statement concerning its description. Moreover, liars have to always save in mind their earlier statements, so that their final deceptive message will appear coherent while retelling their story (Vrij, 2008). Another aspect that increases the mental load of lying is the fact that liars cannot take their credibility for granted as truth-tellers do (e.g., DePaulo, Lindsay, Malone; Muhlenbruck, Charlton & Cooper, 2003; Kassin, 2005). The reasons behind it are at least two. Primarily the liar's stakes (namely the negative consequence of being caught or the positive consequences of getting away) are sometimes higher compared to truth tellers' ones. Secondly, having to strongly account for their "well-acting" and monitoring their demeanour may be as well cognitively demanding for liars at times (DePaulo & Kirkendol, 1989).

Even lies' features have a strong impact on the amount of cognitive load experienced by the senders. For instance, as anticipated, high-stakes lies required an increased effort compared to low-stakes ones; the outcomes of high-stake lies usually really matter to liars indeed, leading them to experience a bigger cognitive load. In addition, stakes become more demanding in presence of self-oriented lies (i.e., lies directly serving the liar in a way that his/her own interest are maximised when truthful statements fail to do so; Cantarero, Van Tilburg & Szarota, 2018), and in turn, the higher negative consequences that are associated with self-oriented lies translate into experiencing a wider cognitive demand (Vrij, 2008).

Across the three types of lies described by DePaulo and colleagues (i.e., outright lies, exaggerations, and subtle lies; 1996), the outright lies result being the most cognitive demanding, especially if compared to the simple act of concealing information observed when telling subtle lies. During the occurrence of outright lies, senders need to invent fictional events and facts while remembering and updating the information that has been told, carrying out executive processes involving a deeper intervention of the working memory (Vrij, Mann & Fisher; 2006). Telling an elaborate lie is more cognitively demanding than providing simple "yes" or "no" answers, since there is much more material to fabricate when telling an elaborate lie.

The same reasoning is expected to be observed when participants of the current study had to deal with complex images: it can be assumed to observe a bigger cognitive load in such a task, given that complex images require fabricating a lie while dealing with a lot of different details. In addition, it has to be stressed that lying is more demanding when the lie itself is not well-prepared or rehearsed in advance. During the deceptive task, subjects have a few seconds to view and interpret the image while even starting to produce a description of it. This situation claims a wide amount of cognitive energy because participants have to come up right away with a fictional scenario while hiding any signs suggesting their deceptive behaviour. Again, it is hypothesized that even under this perspective, complex figures' description results being more challenging. Indeed, in a short amount of time subjects have to process more information and manage details (e.g., inhibition of the truth trace) that are way more developed than those characterizing simple pictures.

2. MODERN APPROACHES TO LIE DETECTION: A SPECIFIC FOCUS ON MICRO-EXPRESSION

2.1. A historical review of the first attempts meant at catching lies

Mentioned that deceptive behaviours are a common phenomenon inside human communication, different disciplines and scientific contexts have nowadays shown interest in working on developing increasingly high-performance lie detection techniques. However, as it is reported by Ford (2006), the first lie detection approaches have very ancient roots: the first method meant at proving the truthfulness of a statement told by an accused subject was documented in China around 1000 BC. The person suspected of lying was at first requested to fill his/her mouth with dry rice and, just after a while, he or she was asked to spit out the rice itself. At this stage, if the rice was found to be dry, the subject was indicated as guilty. The reason behind this methodology relied on the physiological assumption that experiencing fear or anxiety is accompanied by decreased salivation and dry mouth. Nowadays, the works of contemporary authors (e.g., Matsumoto, 2009) have highlighted that fear is reflected in an increased heart rate and in the cognitive feeling of hopelessness that is accompanied by the sense of dry mouth. Nevertheless, it is noteworthy to stress that this symptomatology is not exclusively correlated to fear and anxiety: indeed, even disorders such as depression or panic disorder has similar manifestation, meaning that dry mouth cannot be used as a precise hint for anxiety and fear (and so neither for deceptive behaviour). The lack of specificity and scientific-based evidence led to the execution of the majority of prisoners, regardless of whether they were guilty or not (Vicianova, 2015).

Alongside this first reported methodology, other early techniques failed in showing scientific-based procedures. For instance, in various European countries, a technique known as "the judgment of God" was commonly applied: an accused person claiming the truthfulness of his/her statement was requested to go through a specific act. Based on the favourable or unfavourable outcomes of this act, the claim was labelled as true or false and the accused person was condemned or not. The rationale behind it relied on the belief that God would not let an innocent man suffer unfairly (e.g., Apfel, 2001; Sullivan, 2001).

It was only in the second half of the 18^{th} century that lying (together with delinquent behaviour) became the subject of scientific study, thanks to the pioneering studies of Franz Joseph Gall. The central idea was the existence of an association between different abilities and skull shape; in particular, he perceived the brain as the central organ underlying our mental abilities. Thus, the more active parts of the brain can be easily recognized from the contour of the skull, since Gall believed that these areas were more convex or concave. Mapping the human skull rapidly became a novel scientific discipline that eventually was named phrenology, and its purpose was to look for hints in the skull suggesting the psychological faculties and traits of characters (Rafter, 2005). Phrenology's studies had particular success when applied to legal disputes to expose which part was lying and, in the field of criminology, it played a key role in spreading the belief that delinquent behaviour and lying had to be matters of scientific study. In addition, even if the scientific reasoning was not entirely corroborated, the work by Gall strengthened the medical idea that the brain of some offenders may be affected by pathological alterations. These concepts intervened in reassessing a multitude of sentences preventing mentally ill people from being unfairly sentenced on more than one occasion (Trovillo, 1939, Vicianova, 2015).

In parallel to phrenology, graphology started to be considered a useful and scientific method of lie detection. Its achievements were based on the observations of the peculiarities of handwriting that were interpreted as indicators of personality traits (Schönfeld, 2007). However, early on after its spread, graphology was no more acknowledged as an appropriate tool for lie detection. On the contrary, it was deemed an appropriate procedure for uncovering the authenticity of documents (Vicianova, 2015).

2.1.1. The polygraph's methodology

Soon after the spread of phrenology studies, in 1881 the first modern-like lie detection device called Lombroso's Glove was developed by Cesare Lombroso, who measured the blood pressure of accused persons and record it through a graph or a chart. Originally the methodology was far from embracing a rigorous scientific-based approach, but improvements followed during the First World War thanks to William M. Marston: it reached its final shape in 1921 when this device was exploited to record changes in blood pressure and in breathing while a potential offender was giving testimony (Trovillo, 1939).

Finally, in the 1930s Larson and Keeler worked on and developed a device called "Cardio-Pneumo Psychopgraph" which is now known as the polygraph. In the first place, this polygraph recorded blood pressure changes, respiratory rate, and changes in galvanic skin response (i.e., bioelectric reactivity of the skin). Conversely, modern polygraph analyses rely on recent understandings that pointed out the ratio between thoracic and diaphragmatic breathing as a sensitive indicator of distress and emotional changes (Lewis & Cuppari, 2009). All these bodily changes are measured by the polygraph through the involvement of skin conductance, blood pressure, heart rate and respiration. Notably, the above somatic changes are peculiar to states other than lying, making the polygraph's conclusions prone to be misleading and not being exclusively markers of deceptive behaviours (Brewer & Williams, 2005).

Polygraph's investigation of lies is usually conducted under the shape of two main types of test-question: the Control Questions Test (CQT) and the Guilty Knowledge Test (GKT). The standard is resembled by CQT which always asks the suspect two types of questions: Control questions are usually accompanied by the so-called Relevant question. The control questions concern the subject under examination, but notably are not directly related to the crime; they are intended to induce arousal and provoke embarrassment and make it difficult, to tell the truth. Relevant questions instead concern aspects unmistakably related to the suspect's crime. Finally, the physiological response to both sets of questions is compared (Brewer & Williams, 2005).

The second relevant and frequently used test in polygraphy is the GKT. The guilty knowledge is the focus of the investigation, and it is assumed to be acknowledged only by the actual offender. All the suspects are presented with a series of similar questions that differ in one but yet key aspect resembling the guilty knowledge (e.g., within a murder investigation the GKT questions may all concern the murder weapon while just one will be the actual one). Innocent suspects should be unaware of the guilty detail, meaning that only the perpetrator has reasons to feel threatened by the question involving the key: he or she is expected to react more strongly to the guilty trace than any others (Lewis & Cuppari, 2009).

The centrality of guilty suspects' emotional experience in polygraph examinations emerges. Hints of nervousness, fear and emotional upset are expected to differentiate a guilty subject from an innocent one. However, the source of concerns regarding the rigour of this methodology comes from the evidence that even an innocent suspect might be just as nervous as a guilty suspect. For instance, the innocent subject could fear the test results suggesting his or her guilt, to the point to consider them as probable outcomes. In this regard, the use of a pre-test stands as an effective countermeasure (Bartol & Bartol, 2004): the initial interviews with the polygraph examiner are meant indeed for establishing rapport with the examined and discussing all the procedures. The innocent subject may feel relieved by the understanding that the procedure is effective in revealing the truth, making it less likely that he or she will show bodily changes (Lewis & Cuppari, 2009).

Briefly, it would be misleading to consider the polygraph as a comprehensive lie detection procedure (Lewis & Cuppari, 2009): the polygraph gives back in fact measures of physiological responses that overall are the result of many factors that must be carefully examined (e.g., examiner training, population tested, and techniques implemented). None of the responses identifiable through the polygraph are unique for deception, and neither are they always there when deception occurs (Vicianova, 2015).

2.1.2. Actual cues to deception

The traditional polygraph methodology has played a key role in framing lie detection procedures throughout the past decades, establishing a benchmark for all subsequent efforts. However, it is now clear that the physiological cues signalling increasing anxiety or general arousal are not foolproof in bringing all the information needed for accurate lie detection (Burgoon, 2019; Denault & Dunbar, 2019).

Even though we do not know how to properly use them, some and more specific (compared to physiological ones) cues to deception are available. Getting to know the most reliable deception hints is an intriguing question for researchers, law enforcement and even for lying-tellers. In a review by Sterngalnz and colleagues (2019) all the meta-analyses addressing these issues were collected, with a specific concern about how strongly it is possible to effectively find distinguishing features between lies and truths (Sternglanz, Morris, Marley Morrow & Braverman, 2019).

Researchers usually group deception indicators in nonverbal and para-verbal cues, and content-related cues (i.e., verbal cues). Paraverbal cues are vocal cues that accompany speech behaviour, such as pitch pauses (both filled and unfilled), response latencies or speech errors. Beyond paraverbal cues, in the auditory channel are also included verbal content cues that entail the use of particular word classes, the immediacy or logical consistencies of a statement, and those types of details concerning verbal features. On the contrary, nonverbal cues are mostly related to the visual domain and can be observed in ongoing interactions. Cues such as eye contact, head or hand movements, and leg and foot movements are usually ascribed to nonverbal cues (Sporer & Schwandt, 2006). Finally, differently from paraverbal cues that include aspects such the voice tone, verbal cues or content-related cues are precise characteristics of the wording the senders use, as well as general impressions perceivers of the speaker (Sternglanz, Morris, Morrow & Braverman, 2019).

As concerns paraverbal cues, in a meta-analysis by DePaulo (2003), vocal displays of tensions and nervousness are indicated as the most reliable signs of deception (with an effect size d of at least .20): compared with truth-tellers, liars were observed exhibiting more vocal tension, speak in a higher pitch and appeared more tense and nervous (DePaulo et al., 2003). Sporer and Schwandt went in greater depth within a subsequent meta-analysis (2006) in which they analysed para-verbal behaviours as well. In particular, they observed that liars use to speak in a higher pitch and took longer to begin responding to questions. Nonetheless, they found it evident that moderators affecting the association of deception/cues exist. For instance, liars show higher vocal pitch and response latency only when their speech has to deal with feelings and not exclusively with facts. In addition, researchers noted that the type of experimental setting (different across multiple studies), sender's preparation to lie and degree of motivation has a negative influence on the association of para-verbal cues to deception (Sporer & Schwandt, 2006).

Sporer and Schwandt (2007) conducted an examination of 11 nonverbal signs. Only three of these were found to be significantly associated with deception, showing no correlation at all with truth-telling: nodding, hand movements and foot/leg movements. Going against common sense (Global Deception Research Team, 2006), there was not a statistically significant association between deception and avoiding one's gaze. Moreover, similar to para-verbal cues, the resulting associations are heterogeneous and subject to changes due to contextual and uncontrollable features. In this regard, the content of the lie, the amount of motivation of liars, whether the senders prepared or not their statements, the experimental designs, and the operationalization of the behaviours were found to affect the significance of the single cues (Sporer & Schwandt, 2007).

The specific wording the sender use, as well as the general impression perceivers, has on the sender crucially help differentiate between truth and lies. DePaulo and colleagues (2003) observed that deceptive people usually show less verbal and vocal "immediacy", that is displaying signs of being clear and direct. In their research, liars were labelled as less emotionally involved and seemed more uncertain while reporting their statements (statements were less plausible, less logical and internally discrepant or ambivalent). Liars are not as compelling speakers as much as truth-tellers are: indeed, they used fewer details and showed more complaints leading the experimental receivers to get a negative impression of them, compared to truth-tellers (DePaulo et al., 2003).

2.2. Facial expressions and the leakage of real emotions

Benefits for the advancement of lie detection techniques have come up from research on human facial expressions. It happened well before Ekman that someone addressed the study of facial expressions, driven by their prominent and acknowledged relationship with human behaviour and emotions. In particular, the scientific study of facial expression began with Charles Darwin's "The Expression of Emotions in Man and Animals" (1872). Inside this framework, Darwin has been a pioneer in proposing that some expressions might occur strictly in presence of particular emotions. He was truly convinced that facial movements (of expression) reveal the thoughts and intentions of others more truthfully than words do, suggesting that true and concealed feelings may be shown on some occasions despite effective efforts to hide them. Among other ideas, he originally proposed (then elaborated by other researchers, e.g., Ekman, 1994; Izard, 1994) the theory that some facial displays of emotion are biologically encoded, wired, produce involuntarily and have similar meanings across different cultures: approximately six to nine basic human emotions have emerged throughout the years as universally expressed, each with its own related facial expression. These emotional states include anger, contempt, disgust, fear, enjoyment, sadness, and surprise; in addition, some systems include even interest (Izard, 1977), and embarrassment (Keltner, 1995). Secondly and more surprisingly, Darwin even acknowledged the existence of muscles standing out of voluntary control that might unleash themselves from the efforts to inhibit or mask expressions. These muscles were proposed to reveal true feelings through highly expressive actions (Darwin, 1872; p.54).

As Ekman stressed (2003), these Darwin's ideas are grounded on the foundation of the so-called "inhibition hypothesis" that suggests that if you cannot make an action voluntarily, then you will not be able to prevent it when out-of-control processes like emotion trigger it (Ekman, 2003). To explain this two-process breakdown, it is enough to envisage the neuronal pathways that innervate the face and allows it to contract or relax muscles. In particular, the motor cortex is responsible for those impulses resulting from voluntary effort and meant to elicit a facial expression: the signal is directed towards the facial nucleus (i.e., a cluster of neurons located in the brainstem that belongs to the cranial nerve VII, or facial nerve) that in turn is responsible for the onset of the movements. On the other side, the facial nucleus received input from other lower areas when emotions are aroused involuntarily. Thus, each type of expression may depend upon different independent neural pathways, as confirmed by neuropsychological clinical reports: lesions in the pyramidal systems impair the ability to perform facial movements on request, but leave emotional expression intact as evidenced by the possibility of this type of patients to smile if amused by a joke. Involuntary and emotional facial actions originate in the subcortical areas of the brain and are otherwise driven by the extrapyramidal motor system (Kahn, 1966; Meihlke, 1973; Myers, 1976; Tschiassny, 1953).

The second implication of Darwin's inhibition hypotheses goes beyond the simple distinction between voluntary and involuntary facial muscular actions. More crucial for our deception detection discussion is the notion that if you cannot voluntarily activate a muscle, then you will not be able to consciously inhibit its involuntarily flow whilst a spontaneous emotional expression (Ekman, 2003). Ekman and colleagues questioned this assumption by identifying which facial movements are difficult to move deliberately (Ekman, Roper & Hager, 1980). They succeed in eventually identifying groups of actions that fewer than 25% of their experimental subjects could voluntarily produce (Figure 2.1): by examining videotapes of people being deceptive or telling the truth they were able to spot instances in which the activity of these muscles is not inhibited. Ekman would have later referred to these muscles as "reliable muscles" because their activation should provide leakages of only true emotions, given that they avoid voluntary control (differently from those muscles that most people can easily contract; Ekman, 2003).

Latin Name	Name in FACS	Associated Emotion
Orbicularis oris	24: lip pressor	anger
Triangularis	15: lip corner depressor	sadness
Depressor labii inferioris	16: lower lip depressor	disgust, sadness
Frontalis, pars medialis	1: inner brow raiser	sadness
Frontalis, pars lateralis	2: outer brow	—
(Corrugator = AU 4)	1+4	sadness
	1+2+4	fear
Risorius	20: lip stretcher	fear
Orbicularis oculi, pars lateralis	6: raises cheeks, narrows eyes	enjoyment, sadness

Figure 2.1. *Examples of facial muscles that for most people are out of voluntarily control are reported within the table. Each muscle is associated with a specific expression following the framework of the "Facial Action Coding System" (FACS; Ekman & Friesen, 1978). In turn, each facial expression is associated with an emotion (Ekman, 2003).*

The logic that facial actions that are difficult to make voluntarily should leak concealed emotions is logically correlated to the observation formerly made by Duchenne (1862). He proposed how to distinguish a smile of enjoyment from a non-enjoyment one, by comparing smiles produced through electrical stimulation of the zygomatic major muscle, and smiles generated after a man heard and enjoyed a joke (Figure 2.2): despite both expressions can be labelled as smiles, only the latter entails the pure act of enjoyment that classically it is expected to feel when smiling. Again, both smiles include activation of the zygomatic major, but in addition, the smile elicited when a joke is told is accompanied by the orbicularis oculi muscle's trigger (Duchenne, 1862).

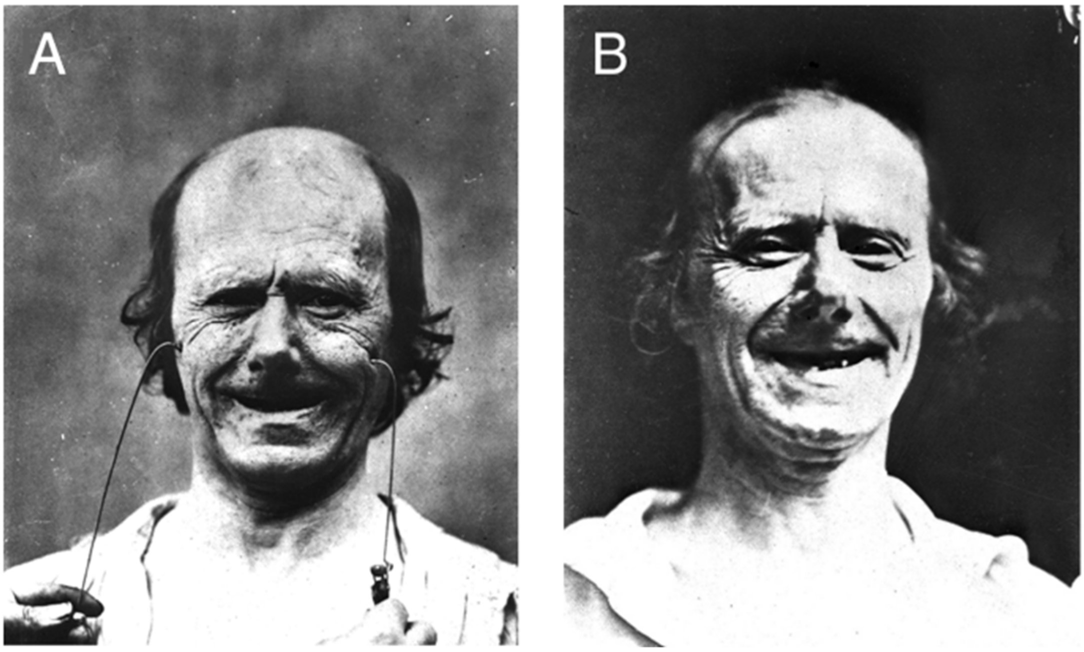


Figure 2.2. *Photographs were taken by Duchenne himself while studying the differences between a smile of pure joy and one "artificially" elicited. In photos (A) the smile was produced by electrically stimulating the zygomatic major muscles. In photo (B) the smile was generated after the subject was told a joke. The crucial aspect to note is the exclusive activation of the orbicularis oculi muscles in figure (B). The best hint that this muscle is not activated is the failure of the eyebrows to lower slightly (Darwin, 1872).*

In agreement with Duchenne, Ekman and colleagues found that just a few people can deliberately contract the orbicularis oculi, stressing the idea that Duchenne's smile (so it is called the pure smile after the French neurologist) might underlie the experience of joy. Consistently, they observed that Duchenne's smiles occurred more often when people were watching an amusing film rather than gory films (Ekman, Friesen & Ancoli, 1980).

As shown by the specific features of Duchenne's smiles, not all of the muscles that produce facial expressions are equally easy to control; there are muscle movements belonging to the face that very few people can make deliberately. Thus, some muscles have been defined by Ekman (1992) as reliable, meaning that they are not available for use in false expression: it is indeed impossible for almost every liar to gain access to them, being out of voluntary control. However, although they cannot be deployed while trying to produce a deceptive feeling, these same reliable muscles are active when the individual feels emotions such as sadness or enjoyment. It is for this same reason that Ekman claims about the need to look for the activity of reliable muscles when dealing with someone that is assumed of concealing his or her emotional state (Ekman, 1992).

2.2.1. Facial Action Coding System

Research has massively looked for a comprehensive method enabling one to predict the presence of emotions within an individual's facial appearances, independently by the employment of machine learning techniques. Among all these tools, the Facial Action Coding System (FACS) by Ekman and Friesen (1978) was found to be the most psychometrically rigorous and above all, widely used (Ekman, 2003; Cohn, Ambadar & Ekman, 2007). FACS theoretically categorizes and measures all the muscular movements that can be observed within the face.

Ekman and Friesen trained themselves in order to isolate their facial muscle: by using EMG needles they wanted to be assured to have included every possible facial movement (and muscles, Figure 2.3) in their system. Only thereafter they resolve to stimulate individual muscles to learn to control them voluntarily (Ekman & Rosenberg, 2005). The selection of facial units was then completed thanks to external observers that were asked to notify up to which point they were capable to spot changes resulting from the various muscles: if two facial changes could not be reliably distinguished, they were combined, regardless of different muscles being involved (Cohn, Ambadar & Ekman, 2007).

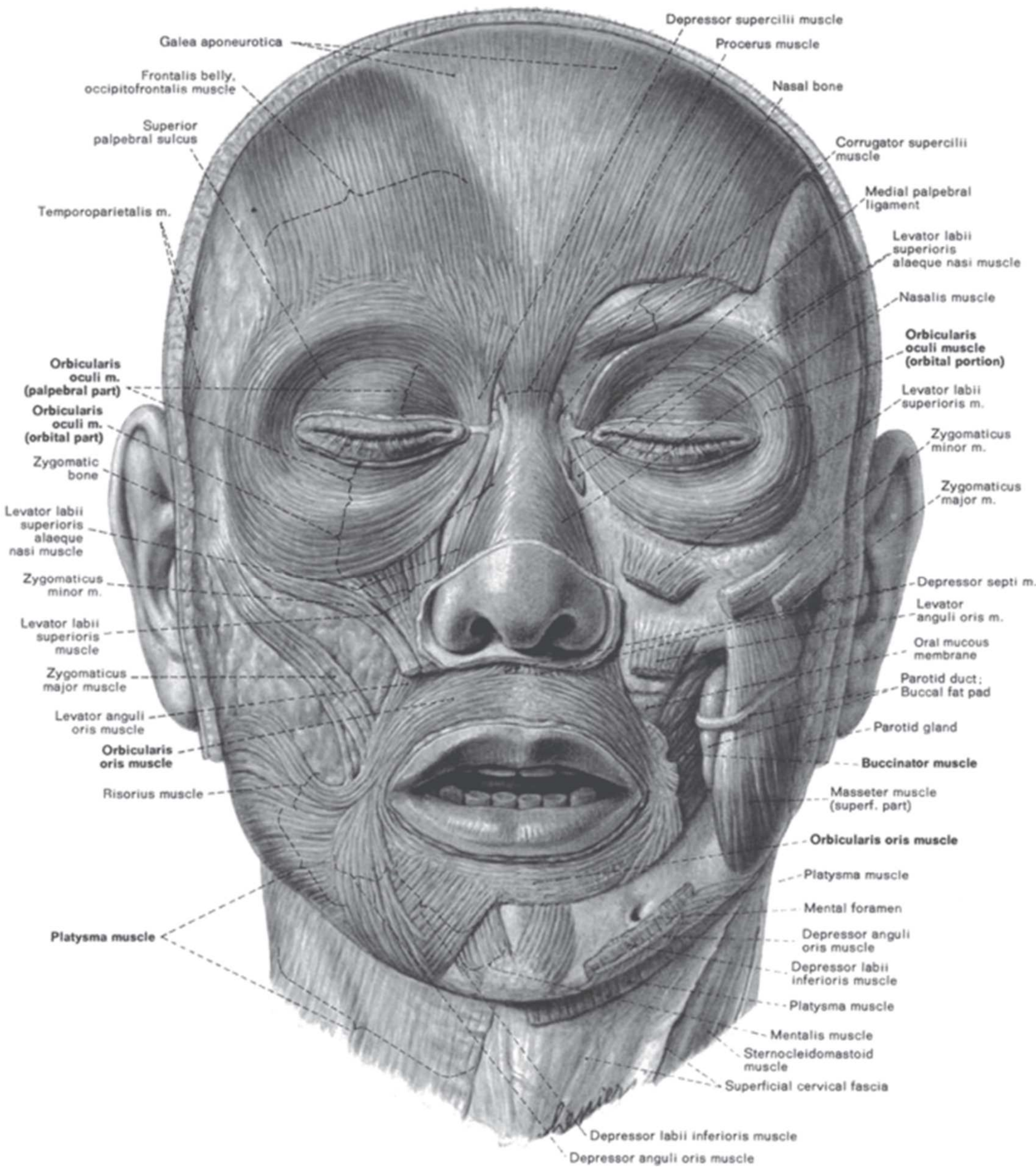


Figure 2.3. *Muscles that are responsible for facial movements are depicted (Clemente, 1997).*

Notably, even if facial events can be described under emotion and non-emotion categories, the FACS (Ekman & Friesen, 1978; Ekman, Friesen & Hager, 2002) does not claim of being an emotional-decoder system. Rather, it is an anatomically-based system for measuring all visually discernible facial movements, on grounds of 44 unique action units (AUs, Figure 2.3), as well as several classes of head and eye positions and movements (Ekman & Rosenberg, 2005). However, "emotion dictionaries" such as the FACS/EMFACS exist in the shape of a computer program for determining whether each facial event includes core movements characterizing specific expressions of emotion (Ekman & Rosenberg, 2005). Unlike systems that employ emotion labels to describe expression, FACS openly distinguish between facial actions and the inferences that can be drawn upon them. Solely in light of a variety of related but yet external resources, such as the "FACS interpretative database" (Ekman, Rosenberg & Hager, 1998), an emotion-based conclusion can be done. Thus, although remaining an inferential step extrinsic to FACS, it is possible to combine each AU with an emotion-specified expression (Cohn, Ambadar & Ekman, 2007).

Every single AU is labelled by a numeric code, the designation of which is completely arbitrary. There is not a 1:1 correspondence between muscle groups and AUs: muscles can indeed activate and contract in different ways and directions producing a really broad group of actions. An example is the frontalis muscle which contraction of the medial portion raises the inner corner of the eyebrows (i.e., AU 1), while contraction of its lateral portion raises the outer brow (i.e., AU 2; Ekman & Rosenberg, 2005).

FACS entails 9 AUs in the upper face (Figure 2.4) and 18 in the lower face, with the addition of 14 head positions and movements, 9 eyes positions and movements, 5 miscellaneous AUs, 9 action descriptors (i.e., movements lacking of anatomical bases), 9 gross behaviours and 5 visibility codes. Apart from some exceptions, these AUs are organized by region of the face and each has both a numeric and a verbal label. AUs from the head and eye positions, amongst others, not rarely are omitted in FACS scoring, regardless of growing evidence about their relevance to the interpretation of facial expression and emotions. It is common that similar facial movements, such as smiling (AU 12), vary in meaning coherently with their temporal coordination with head motion. By way of example, in embarrassment smile intensity increases as the head moves forward, whereas a decrease in intensity when the head moves back toward frontal orientation (i.e., negative correlation, Cohn et al., 2004, Keltner & Buswell, 1997).

Even combinations of AUs are worth considering. They occur as additive or non-additive, depending on whether the appearance of each AU is independent: in non-additive combinations, AUs modify each other's appearance when they are held simultaneously. To make it clearer, non-additive combinations are similar to co-articulation effects happening in speech, in which one phoneme modifies the perceived sound of those with whom it is concurrent. This is what takes place when AUs 1 + 4 occur in sadness (Darwin, 1872). AU1 alone is equal to inner eyebrows pulling upward, while in AU4 alone eyebrows are pulled together and downward. Conversely, when these two AUs are concurrently AU4 modifies by presenting itself as eyebrows still pulling together but rising: the final combination translates to an oblique shape of the brow that causes horizontal wrinkles to emerge in the forehead's centre (Cohn, Ambadar & Ekman, 2007).

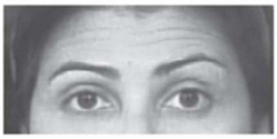
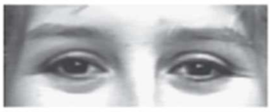
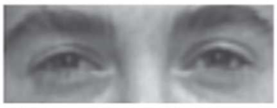
AU	Description	Facial muscle	Example image	Interrater agreement (Kappa coefficient) (tolerance window in seconds)			
				1/30th	1/6th	1/3rd	1/2
1	Inner Brow Raiser	*Frontalis, pars medialis*		.73	.79	.81	.83
2	Outer Brow Raiser	*Frontalis, pars lateralis*		.66	.71	.74	.76
4	Brow Lowerer	*Corrugator supercilii, Depressor supercilii*		.58	.64	.67	.70
5	Upper Lid Raiser	*Levator palpebrae superioris*		.68	.76	.79	.82
6	Cheek Raiser	*Orbicularis oculi, pars orbitalis*		.72	.78	.82	.85
7	Lid Tightener	*Orbicularis oculi, pars palpebralis*		.44	.49	.53	.56
9	Nose Wrinkler	*Levator labii superioris alaquae nasi*		.67	.76	.81	.83

Figure 2.4. *Action units of the facial action coding system belonging to the upper face are reported (Ekman & Friesen, 1978). In addition, it is reported the associated interrater agreement for each AU. They are calculated using coefficient kappa (Sayette et al.,2001) which controls for chance agreement. (adjusted from Cohn, Ambadar & Ekman, 2007).*

Crucial within FACS' framework is the scoring of AUs. Different degrees of freedom exist concerning the level of detail with which AU coding is performed. Generally, a major coding dichotomy is considered between the mutually exclusive coding and the combined one. In selective coding, only predetermined AUs are coded, while others are completely ignored. Conversely, in comprehensive coding, each AU (found in a specific and chosen segment of facial behaviour) is considered. Both approaches have advantages and disadvantages, and usually, the decision of which one to choose requires the precise examination of the research question. For example, in research that is intended to understand whether specific facial patterns standing for embarrassment exist, it might be needed to comprehensively code video frames of facial behaviours during which subjects report embarrassment (Cohn, Ambadar & Ekman, 2007). Today it is known that embarrassment is associated with a particular AUs' sequence (i.e., AU 12 followed by AU 24 and then AU 51 + AU 54); considering just a subset of AUs would have rendered impossible the discovery of this pattern (Keltner, 1995).

Regardless of employing selective or comprehensive coding, examiners can determine the level of detail of each AU. Leaving aside the rudimentary coding answering just if an AU is present or absent, coders usually pay attention to the intensity or strength values of the actions. The value scale is classically divided into five levels of intensity coding (A, B, C, D and E) where A is the least intense movement (a trace) and E is the maximum strength one. A drawback of this coding is the subjectivity of the processes, that is the guidelines for intensity scores rely on the examiner's assessment criteria: the issue becomes particularly relevant for mid-range intensities because it is required major effort to evaluate and establish acceptable levels of reliability (Cohn, Ambadar & Ekman, 2007).

Moreover, AU scores cannot be always quantified in singular terms. AUs are often pre-coded into defined patterns and in some instances can independently linger and merge in the background. This evidence raises the occasional need of considering a set of singular AUs as a whole while scoring them. In doing so, a group of AUs overlapping in time or appearing to define a perceptually meaningful unit of facial movements can be considered with a view of a single display (or as an event, Oster, 2001). This approach is adequate from the perspective that facial behaviours do not occur in a continuous fashion but rather as events that typically show themselves as discrete episodes. AUs that take place together are related to some extent and form an event (Oster, 2001; Oster et al., 1996).

Finally, researchers do not always agree on how to define an event. One solution usually adopted is to define as events those combinations of AUs that are known to co-occur: remarkably co-occurrence rates are lacking and often are population-specific, making it difficult to find coherent data. Secondly, events are often involved in those studies that look for AU combinations that are commonly associated with an emotion, even though this approach is potentially prone to violate the logic underlying FACS, which is to keep description separate from inference: works that have used event coding in such a way have typically failed in reporting the basis on which they define an event, ending up to rely more on judgments than on an evidence-based approach (Cohn, Ambadar & Ekman, 2007).

2.2.2. What are and why study micro-expressions when dealing with deception detection?

Reliable muscles' importance for deception detection can be easily ascribed to the early discovery of micro facial expressions. Ekman alongside Friesen (1969) found out for the first time about their existence while examining a couple of films involving statements of their psychiatric patients. The examination of a singular clinical interview paved the way for the following understanding: Ekman used a slow-motion projection to examine an interview of an old patient that successfully concealed her intention to commit suicide. In freeze-frame, the true emotion of anguish came out in the form of a brief-expression barely perceptible to the untrained observer. Ekman and Friesen assumed that this micro-expression, even if subtle, had to be a sign of repressed or deliberately suppressed emotions (Ekman & Friesen, 1969). As anticipated, the theory behind micro-expressions was then developed by positing that when people try to mask their true emotional state, expressions coherent with their real state will emerge briefly on their faces. Some facial muscles (i.e., the reliable muscles) usually avoid voluntarily control, and some automatic and uncontrollable displays of emotion will produce briefly detectable "leakage" (or micro-expression; Ekman, 1985; Jordan, Brimbal, Wallace, Kassin Hartwig & Street, 2019). These micro-displays may differ in their nature, ranging between fragments of a squelched, neutralized, or masked display and full muscular movements associated with macroscopic expression of affection (Ekman & Friesen, 1969). Coherent across all micro-expressions is instead the span: they flash on and off within the face's muscles in less than one-quarter of a second. Additionally, to their short-lasting span, micros prevent untrained people from being able to detect them (Ekman, 1985).

After stating their importance, it is noteworthy to stress that understanding the nature and meaning of micro-expressions might be controversial at times: micros can emerge in cases of both voluntary concealment and unconscious emotion repression, and no features intervene and help us to properly categorize them. Ekman himself stressed that micros do look the same in both concealment and repression (Ekman, 2009). Otherwise, the context in which the micros occur is the only significant means to catch the meaning behind the same micro-expression. Assuming that micro-expressions arise in a subject's face while being examined on his or her statements, 4 contextual aspects have to be considered. First, the nature of the conversational exchange (1) during which the subtle expression emerges has to be examined, as well as the conversational characteristics (2) (e.g., it is a first meeting, a casual conversation or a formal interview etc.) and the information about what has transpired in this conversation up to know (Ekman, 2009). A third issue is the speaker turn (3), which is the exact moment in which the micro shows up: it may occur respectively when the subject being examined is speaking or when he or she is listening. Finally, the last contextual feature is congruence (4): it resembles whether the emotion displayed by the micro-expression fits or contradicts the speaker's simultaneous statement topic, voice tone, gestures, and postures. The same reasoning applies when the person is listening, but then the focus shall be on the evaluator's demeanour (Ekman, 2009).

All the addressed contextual aspects have also an influence on the potentiality of detecting a lie through the mean of micro-expressions, although remarkably it has been mentioned that micro-expressions are nothing more than signs of concealed or subtle emotional behaviour. Conversely, situations in which the aid of a lie detection system would be beneficial are most of the time unrelated to emotional experiences: for instance, in criminal settings, a potential lie would hardly entail what emotion is being felt at the moment, but rather it might regard an action. Nevertheless, although this theoretical gap between contexts in which deception is frequent and micros is big, it is not misguiding to admit that emotions can become involved in lying processes regardless of the topic of the deceptive statement.

2.2.3. Evidence and pitfalls in the relationship between micro-expressions and deception

Almost every model of lie detection agrees on recognizing that cues to deceit are caused by not only cognitive factors but also by emotional factors, such as guilt or distress, fear, or even emerging experiences of enjoyment (Zuckerman et al., 1981). Reasons behind the experience of these emotions are proposed to take place when the liar feels fear of getting caught, distress or guilt at telling the lie, or even contempt or disgust for the target of the lie (Frank & Steven, 2013). Thus, to the extent that the deceptive processes generate emotions, it is plausible to predict that signs of the aforementioned emotions may betray a lie (Frank & Ekman, 1997). Prototypical situations in which a liar might be emotionally-loaded are those in which the stakes are really high for getting away with it. Unfortunately, studies that aimed to verify the existence of these emotional-behavioural cues have commonly overlooked the need of including high-stakes lies. It is as such that the most comprehensive meta-analyses concerning behavioural cues (DePaulo et al., 2003) have failed to find a significant effect size for some (crucially, not all) facial clues of lying (Frank & Svetieva, 2015). Frank and Svetieva (2012) observed that if the high-motivation (a stand-in for stakes) studies were separated from the others, a stronger effect for emotion-based cues was evident.

Several studies that employed systems for the taxonomy of human facial movements, such as the Facial Action Coding System (FACS), have succeeded in showing that facial expressions of emotion can suggest deception. A significant portion of these facial cues have been classified as micro-expressions and that is particularly true for studies featuring individuals lying about their feelings. An example comes from Frank and colleagues (2014): they asked participants (i.e., members of politically active groups) to choose whether to take a $100 check made out to one of their rival parties. Participants were told that they were going to be interviewed by trained and expert subjects in the field of lie detection: getting caught lying (denying oh having earned anything) while accepting to take the money was translated into a monetary loss for their group, and in a mutual increase for their rivals. Conversely, in case they lie and get away with it, they were rewarded with even more money. Eventually, the results showed that 72% of the 132 participants could be rightly classified as being deceptive or honest by the presence or absence of negative emotional experiences such as fear, distress contempt or disgust. Furthermore, among those emotions that betrayed lying statements, 51% lasted 0.5 s or less, while 30% of them were even less than 0.25 s.

A final note that may stand as a source of concern is that an equal rate of micro-expressions occurred in those truth-tellers that showed these same negative emotions (Frank et al., 2014). As such, they were classified under the light of false positives, proving right other studies' conclusions that highlighted that micro-expressions might occur even in honest scenarios (Porter et al., 2012).

However, Frank and colleagues were able to further demonstrate how these subtle expressions act of pure involuntary nature. By the mean of the last questionnaire, participants were asked about which strategies they used to fool the interviewer: those liars that stated to have paid attention to their face's appearance by using the strategy for managing their facial expression showed exactly the same rate of negative-betraying emotions as those who did not report having tried to deliberately manage their expressions. Conversely, truth-tellers that stated they employed a face-monitoring strategy showed a significantly less amount of these emotional expressions (18%) than those who did not report using such a strategy (35%). These results were interpreted in the light of a hard time that liars may experience when concealing micro-expressions: this aspect was further questioned by specifically instructing lying subjects to hide their expressions of fear and happiness while being interrogated. Although participants succeed in decreasing the duration and intensity of their facial expressions, yet almost all of them showed signs of these emotions (Frank et al., 2014, Hurley & Frank, 2011).

The idea discussed earlier concerning the occurrence of micro-expressions even within honest statements highlights how easily might be to mistakenly draw a conclusion upon the onset of these short-lasting expressions. Ekman himself addressed the issue by mentioning the so-called Othello's error, who accidentally assumed that Desdemona's expression of fear suggests betrayal: Othello, as everybody who falls in the error named after him, failed to understand that emotions do not tell you their cause. The fear of being disbelieved looks the same as the one experienced when lying, and it is commonly experienced even by truth-tellers when presented with a high-stake interview (Ekman, 2003). Only through further questioning, it is possible to realize whether the concealed fear is the result of feeling extremely under examination or concerns the anguish of being caught. As a result, behavioural cues such as micro-expressions can be viewed in terms of "hot spot", rather than signs of lying: they mark where it is needed to investigate through the mean of questioning and background checks. Alternative explanations have to be taken into account regarding why the behaviour occurred so that they might be excluded before highlighting the presence of a lie (Ekman, 2009).

Another and last source of concern comes from the evidence that not everyone who suppresses or represses an emotion shows a micro-expression: in all the research studies conducted by Ekman (e.g., 2009) micros have been observed in about half of the people who voluntarily lied within the experimental settings. If proven right, this conclusion would undermine the existence of differences between genuine and deceptive displays, even if some clarifications have to be posited before evaluating those results.

To seek a solution, it has to be acknowledged that up to now one of the biggest issues related to the use of facial expressions to detect deception might have been the exclusive engagement of an encoder-decoder perspective (Zloteanu, 2020). First of all, under this view poor detection accuracy is only explained in terms of a scarcity of cues available or the inability of an examiner to detect them. Unfortunately, both explanations do not take into account the validated conclusion that some of those muscles defined as "reliable" can actually be activated in the absence of genuine affect too: all the different arguments for liars being unable to produce genuine-looking expression of emotions relies on the Duchenne smile that, in some occasions, has been proved wrong (Zloteanu, 2020). For instance, Krumhuber and Manstead (2009) worked on facial action units' activation in situations where individuals were feeling respectively genuine or faked happiness. Going against their expectations, not all instances of genuine happiness resulted in the facial activation of those assumed reliable muscles. In addition, recently Gunnery and Hall (2014) showed that a small - yet remarkable - amount of individuals can deliberately generate genuine-looking smiles while being no less perceived as more persuasive (Zloteanu, 2020).

In brief, accounting for only an encoder-decoder (i.e., emotion-based) perspective is not enough. It has been highlighted that across all the available facial cues (that are assumed to be there in most but not all, lying contexts) it is necessary to discriminate those that are authentic and may suggest veracity: encoders have been found imprecise while confronting with these tasks (Jordan et al., 2019), suggesting the need of employing new and more rigorous methodologies. Liars are, on many occasions, good and strategic communicators that at times can be capable of supporting their behaviour with deceptive and manufactured emotional cues, rendering ineffective human-based analyses of emotional expressions (Zloteanu, 2020). Therefore, standing as the focus of the current work, machine learning techniques are presented in the next paragraphs as a useful tool, given that they have displayed promising scores performances in detecting whether a subject is lying, both by employing just facial cues or multi-modal approaches.

2.3. Human beings' performance and the need for an artificial detector of micro-expressions

Given that FACS was developed as a tool for measuring the activity of facial muscles, it appears reasonable to employ it in an attempt to extract micro-expressions, and in turn to detect deceits. FACS can effectively be used as a low-level feature extraction tool, standing as a starting point for a human-based detection of micro-expressions (Takalkar, Xu, Wu & Chaczko, 2016). Furthermore, studies supported the notion that people can be trained to detect micro-expressions, and that training usually persists over time (Hurley, 2012). Frank and colleagues (2014) observed that training procedures can intervene and lead to greater than chance-level improvements. Usually, pieces of training cycles are based on the "Micro Expression Training Tool" (METT) that was developed by Ekman (2002) with the purpose of training people to improve accuracy in recognizing micro-expressions. The effectiveness of METT has been corroborated by different studies (e.g., Frank et al., 2014; Matsumoto et al., 2012), showing that training in micro-expressions translates into statistically significant improvements in the abilities of people to detect concealed emotions when presented alongside real-time displays of subtle expressions. Better performances in the ability to detect deception have been claimed too (Frank & Svetieva, 2015).

If on one hand, much literature has proved that the METT is able to increase the recognition of covert facial expressions, on the other it seems too early to view it in terms of aid in detecting deception. Within an experimental paradigm, Jordan and colleagues (2019) made their participants undergo the advanced module of METT: their scope was to evaluate if undergoing this procedure would lead to better outcomes in detecting deception. Experimental participants' scores were then compared with those of a control group and with those of a fictional training-procedure group (i.e., "Interpersonal Perception Task"). The deception detection task was conducted by asking the participants to evaluate videos containing respectively truth and lies statements: the stimuli were collected from different and preceding deception detection studies (Jordan et al., 2019). Overall, the results showed that belonging to the METT group does not affect the scores of detection: the overall accuracy was slightly below the chance level of 50% and no statistically significant differences were found across the scores of the three groups (Figure 2.5), thus not supporting the claim that a training procedure might ameliorate the human being's capacity to detect deception, besides an improvement in the detection of micro-expressions was observed (72% of the participants in the METT condition reached the 80 % threshold of accuracy in detecting micro-expressions) (Jordan et al., 2019).

	Training			
Video set	METT	IPT	No training	Total
Vrij and Mann (2001)	44.80	32.10	36.70	37.90
Street et al. (2011)	53.50	63.30	60.00	58.90
Sorochinski et al. (2014)	43.30	50.00	46.70	46.70
Toomey (2013)	46.70	43.30	50.00	46.70
Kassin et al. (2005)	43.30	46.70	26.70	38.90
Total	46.30	47.30	44.00	45.90

Figure 2.5. *The participants' percentage scores concerning their lie detection accuracy are displayed by the type of training underwent (METT: micro-expression training toll, IPT: interpersonal perception task; No Training: control group) and different video sets (collected from previous studies). The differences between the scores obtained in different sets of videos were explained in terms of content (i.e., relevant to security screening, missing person, and past transgression); speakers; and stakes (Jordan et al., 2019).*

This study by Jordan and colleagues failed to find any correlation between the ability in spotting micro-expressions and deception detection. Other studies that have been already mentioned (e.g., Porter, Brinke & Wallace, 2012; Porter & Ten Brinke, 2008) have come to the same results, claiming that even if micro-expressions may occur, they are rare and emerge both in presence of truth-telling and lying. People (both trained and not trained) have a hard time detecting them because the time duration of micro-expressions is 0.5 s at low intensity (Yan, Wu, Liang, Chen & Fu, 2013). Moreover, highlighting and distinguishing authentic facial cues from non-authentic is another key overlooked aspect of lie detection: micro-expressions do not always present themselves under the same shape, and not rarely people may express them in different ways. Some individuals are capable of contracting muscles that other individuals cannot, suggesting the need to look for subtle details that even trained subjects may not take into account.

It is in this light that in the current work machine learning system have been considered as the most logical solution: artificial systems can analyse facial-expression dynamics at a much higher temporal solution and with a more complex description than it was feasible through a human-based and manual coding procedure (Bartlett, Littlewort, Frank & Lee, 2014). And to that end, studies in the literature have already shown how promising scores machine learning systems may make available for the deception detection field.

2.3.1. How artificial intelligence can be beneficial for lie detection's advancement

The majority of the most up-to-date strategies meant for improving lie detection focus on techniques that involve artificial intelligence (AI) approaches. Within an AI system, the human examiner is often able to play a less visible role, making the evaluation processes autonomous and less likely to reach biased conclusions. AI allows indeed to extract higher-level features, by exploiting non-visible structures in complex input distributions that otherwise would remain undetectable, such as light changes in the pitch of the voice or subtle muscle movements within the face (Oravec, 2022; Bhamare, Katharguppe & Nancy, 2020).

Another major concern of many lie detection procedures that can be addressed through AI is the possibility of liars escaping from being identified thanks to faking and coaching (Alliger & Dwight, 2000): with some of the AI data collection systems, these faking efforts are made more difficult due to the uncertainty of how, when and what data are being investigated. The modalities for acquiring data have tremendously increased far beyond our senses' abilities and now also involve tools that recollect and gain evidence even without the subject's proximity or consent. For instance, this is what happened with instruments like eye scanning (i.e., used to collect eye blinking patterns) or webcams. Other approaches such as fMRI, even though they are invasive, are currently providing complex dataset that requires machine learning to be interpreted, potentially diminishing the transparency of the system involved (La Tona et al., 2020).

Recently, AI researchers have invested resources even in developing "corpora" of training examples to be used in machine learning. It is the case of Takabatake and colleagues (2018): they worked on the creation of a "liar" corpus that collects inside (and set for analysis) various human expressions typical of situations that reportedly involve prevarications. Forms of biases can be included as selected stimuli and skewed in different dimensions, such as detailed racial or gender orientations (Tambe et al., 2019). Unfortunately, issues might still occur; training samples are often created through social media scraping, crowdsourcing, and other processes that undoubtedly can introduce bias in such a way that the developers might not even take into account (Oravec, 2022).

That being stated, maybe the greatest advantage brought by the exploitation of AI systems is developing new lie detection-related constructs that otherwise would be difficult to utilize or even challenge. AI enables the craft of complex constructs such as those related to micro-expressions and deceits biomarkers in general. For what concerns micro-expressions, machine learning capabilities for analysing a large amount of data concerning facial expressions have been constructed with the scope of determining which subtle facial changes and combinations of physical hints are correlated with deception (Oravec, 2022). An early and prototypical effort that tried to involve AI in lie detection by the mean of micro-expressions is the Silent Talker (Kennedy, 2014): it consists of a digital video camera hooked up to a computer that records a subject while he or she is being presented with an interview. The AI involved in the system identifies the non-verbal micro gestures of the interviewee: these are unconscious responses that the Silent Talker picks up to determine whether the subject is lying or not. Thus, AI can intervene in fragments of lie detection analyses by rendering conceivable the collection and observation of cues that otherwise would remain undetectable, making it even more challenging to look for innovative deception-suggesting evidence.

Various directions have been followed within the AI-deception dichotomy. Features coming from videos, audio and text material have all been considered when attempting to work on automatic deception detection. By stating that up to now none of these frameworks has yet revealed itself as entirely satisfying, it is still wise to urge that some research paths are more promising than others.

2.4. Micro-expressions, machine learning and lie detection: a review of the literature

In the following, an exhaustive review of the studies that exploited machine learning to detect deception through the analyses of facial micro-expressions is reported.

Owayjan and colleagues (2012) were among the first to propose a lie detection system that automatically extracts facial micro-expressions. They employed a high-speed camera (acquiring 25 frames per second) to capture participants' faces. The procedure was carried out thanks to a multi-core processor (Embedded Vision System, "EVS"), meant to run the detection algorithm programmed with the NI LabVIEW Operating system. As soon as participant interviews were conducted, the EVS started to convert the videos in the sequence of frames to facilitate the analysis. Secondly, geometric-based dynamic templates were applied to specific parts of the face to mark key features of the expressions. The program started by reading one frame at a time and simultaneously extracting frames by the flow of two parallel loops. In particular, the second loop was meant to process the saved image according to predefined templates (Figure 2.6 a). All the templates were predefined using a Vision Assistant and represented specific areas of the faces; every time one of them was detected the program measured nine different distances (Figure 2.6 b) between crucial points found within the template itself (e.g., the horizontal length of the mouth). The distances were then separately saved in unique arrays. Eventually, the total arrays correlated to all groups of points were compared in accordance with preprogrammed rules derived from facial appearances of emotion (organized through the FACS system).

Figure 2.6. *(**A**, image on the left) reports the template detection on the face of an experimental subject, whereas (**B**, image on the right) reports the distances shown on the same face. In general, Templates represent the following areas: the left and right edges of both eyebrows, the left and*

right edges of the eyes, the left and right edges of the mouth, and the cheeks (Owajyan, Kashour, Haddad, Fadel & Souki, 2012).

The combinations stored in the arrays and obtained by comparing distances between different frames lead to the generation of a basic pattern of expression (i.e., anger, contempt, disgust, fear, happiness, joy, sadness, and surprise): crucially, each basic expression had specific points measurements and distances combination. Finally, the program iterated over the expressions coded in all the arrays and noted the duration of each expression per se: if expressions were repeated less than 5 times consecutively it was marked as micro-expression. Owayjan and colleagues assumed that the system was capable of detecting deception based on whether a micro occurred: when the system was tested on four subjects presented with a questionnaire containing both control and relevant questions the recognition accuracy for lying statements was 85% (Owajyan, Kashour, Haddad, Fadel & Souki, 2012).

Later on, Su and Levine (2014) paved the way for the analysis of real case scenarios: they collected a database (standing as the first-ever made) consisting of high-stake deception video clips from real-world situations (i.e., 324 video clips of people asking for help to find missing relatives or potential murders that killed them). As Owayjan and colleagues, they assumed the existence of emotional leakages of deceits and found a series of AUs that in their opinion were indicators for distinguishing truth-tellers from liars in high-stake situations. The method they proposed consisted in looking for the AUs depicted in (Figure 2.7) to discern deceptive and honest subjects.

Emotion	Truth-Tellers	Liars
Sadness	AU1+4, AU15 (Genuine)	AU1 (Fake), AU2 (Fake), AU1+2 (Fake)
Happiness	NA	AU6+12 (Genuine), AU12 (Fake)

Figure 2.7. *AUs and their combinations are reported that can distinguish between truth-tellers and liars when experiencing sadness and happiness (Su & Levine, 2014).*

Through the Pittsburgh pattern recognition software (PittPatt), three primary facial land markers were noted in every and single video frame: left eye, right eye and nose base. Faces were then spatially normalized, and nine facial areas were located according to an anthropometric face model. All the subsequent analyses were conducted in each region of interest (ROI) for each frame of a unique video. Different methodologies were then applied to collect eye blink processes, eyebrow motion, wrinkles, and mouth motion.

After these procedures, each video was decomposed into nine temporal sequences of specific face regions and for each region, one or two feature vectors (events) were calculated. Finally, each feature corresponded to a facial AU. Noting that some AUs are likely to occur simultaneously, secondary features (i.e., congruence of events) were computed too.

To create a coherent representation of every video, Feature Temporal Volumes (FTVs) were considered and viewed as basilar descriptors of a video clip: taking into account the temporal context surrounding the frames was a key aspect in achieving a compact representation of facial movements.

Having a look at the results, liars were considered as the positive sample and the accuracy of spotting them was accounted as a true positive rate, on the contrary, spotting truth-tellers was accounted as a true negative rate. On average, the accuracy percentages reached 74.52%, posing the first step for an automated attempt in the literature at proving the veracity of facial cues of deception (Su & Levine, 2014).

Bartlett and colleagues (2014) tested both human observers' and computer vision systems' abilities to categorize real and fake emotional expressions of pain. Videos of people experiencing pain were collected inside the experimental paradigm by asking single subjects to submerge their hands in cold water and recording their experiences. Human detectors' scores of real against fake pain did not diverge significantly from chance level: making people follow a training programme just increases slightly the performance by around 60%.

Videos were then presented to a computer vision system called the Computer Expression Recognition Toolbox (CERT). CERT automatically detects frontal faces in the video and computes every frame with respect to a set of continuous dimensions, such as facial muscular actions through the employment of FACS. Unlike manual coding, CERT gives an instantaneous output of facial-movements information, adding even facial-expression intensity and dynamics at temporal resolutions.

Bartlett and colleagues used a pattern-recognition approach (Figure 2.8) to evaluate the detection performance of CERT: all the 60s videos were used as input for the system one at a time. A set of dynamic descriptors was extracted from the output for each of the 20 AUs recognizable by CERT: one set of the descriptors described facial movements event whereas the second one described the intervals between events. Next, a classifier (support vector machine, SVM) was trained to discriminate between real and fake pain using the described descriptors. The SVM combined information from multiple AUs through a sequential feature selection procedure that started from the AU that gave the best individual classification accuracy. Afterwards, AUs that gave the best performance, when combined with the previously selected AUs, were given to the model.

Finally, the computer system was tested on one video at a time: it achieved a detection accuracy of 85%. Conducting a more in-depth analysis it was possible to extract the most informative AUs for differentiating real pain from faked one: mouth opening (AU 26) was the most informative, followed by lip raise (AU 10), lip press (AU 24), and brow lower (AU 4).

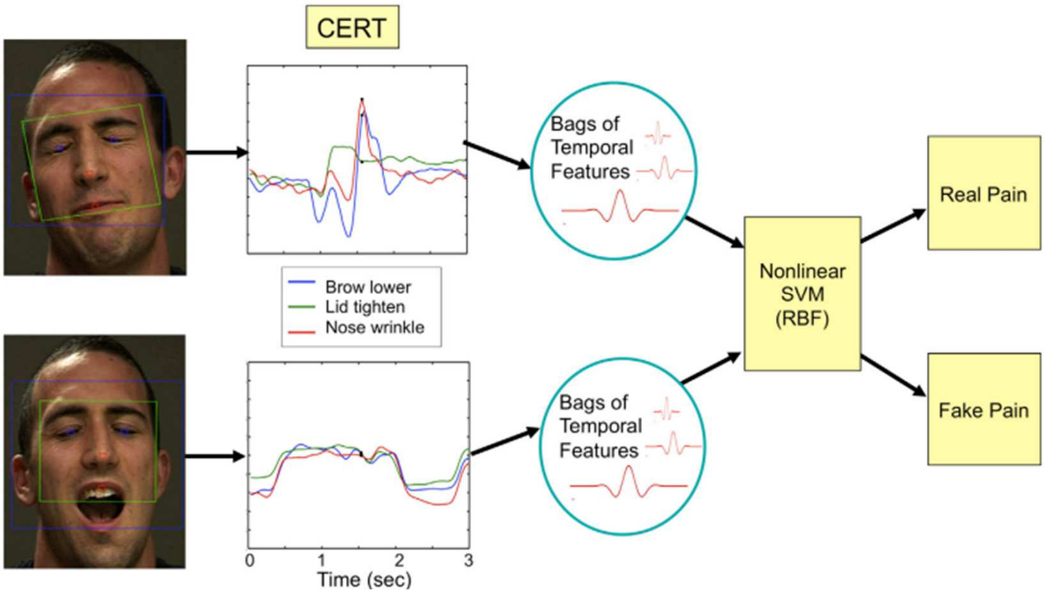


Figure 2.8. *Videos concerning facial movements were processed by the CERT alongside the magnitude of 20 facial actions over time. The CERT's output on top resembles pain, whereas the bottom one is the representation of the same three actions but related to fake pain. The dynamics differ. As a next step, the expression dynamics were measured with a bank of eight temporal Gabor filters and reported in terms of bags of temporal features. These data were given to a non-linear SVM classifier that was planned to categorize real and fake pain. Classification parameters were learned from twenty-four video samples of real/fake pain experiences (Bartlett et al., 2014).*

A subsequent and cutting-edge study was developed by Pérez-Rosas, Abouelenien and Mihalcea (2015). The study was presented by the authors themselves as the first attempt at building a multimodal system for detecting deception: differently from the previous studies, Pérez-Rosas and colleagues resolve to analyse more than one potential cue for deception, underling the importance of considering more than just one evidence when dealing with deception detection. Moreover, the study was conducted on a new database (121 video clips) concerning real-life trial data using text and gesture modalities. The veracity judgments were interpreted in light of the court verdicts.

Both verbal and non-verbal behaviours were analysed and annotated to understand their relationship with deception. Non-verbal features concerned those gestures (facial expressions and hand gestures were the focus) observed during the interactions in the video clips. The gesture annotation was conducted with the MUMIN coding system which is a standard multimodal annotation scheme for interpersonal interactions. Within the MUMIN system, facial displays consist of several facial expressions associated with overall facial expressions, eyebrows, eyes and mouth movements, gaze direction, and head movements too. Furthermore, the multimodal annotation was performed by two annotators using the Elan Software: annotators were asked to identify the facial displays and hand gestures that were most frequently observed during the clip duration. At the end of the annotation, it was possible to derive the non-verbal features that were created when a gesture was observed during the majority of the interaction. On the other hand, verbal features consisted of unigrams and bigrams derived from the bag-of-words representation of the video clips transcripts: features were encoded as words or word pairs frequencies and excluded if their frequencies were lower than 10.

After checking all the observable differences between the deceptive and truthful groups, further examinations were conducted to check the performance of the extracted features with a machine learning approach. Two classification algorithms were used: Decision Tree (DT, combined classifier using all features at once) and Random Forest (RF, individual classifier relying only on facial display features).

Facial displays seemed to contribute the most to the classifier performance, but it was clear too how merging features might be beneficial for increasing detection performances. The final and overall achieved accuracy was in the range of 60 to 75 %, outperforming humans scores that, within this work and through different modalities, reached a maximum accuracy of 51% (Pérez-Rosas, Aboulenien, Mihalcea & Burzo, 2015).

Following the model of Pérez-Rosas et al., Wu and colleagues (2018) developed a model for automated deception detection including features from several modalities. Micro-expressions were included as well, but here it is stressed the importance of analysing them from videos rather than static images: they proposed to use a motion dynamic for recognizing micros, which is making it much easier for a human detector to spot them.

To accomplish this task, the authors designed a two-steps feature representation for capturing dynamic motion signatures: for low-level features representation dense trajectories representing motion were used, whereas for high-level representation a facial micro-expression detector using low-level features was trained. The general framework consisted of three steps: multi-modal feature extractions, feature encoding and classification (Figure 2.9).

The Improved Dense Trajectory (IDT, Wang et al., 2016) was employed for its effective performance in action recognition: IDT analyses local feature-correspondences in subsequent frames and estimates the camera motion employing RANSAC. As a next step, the histogram of oriented gradients, the histogram of optical flow, the motion boundary histogram and trajectory descriptors are computed within the space-time volume. Eventually, the motion boundary histogram seemed to be the best among the other descriptors, given that it captures derivatives of motion rather than first-order motion information: Wu and colleagues realized that since they wanted to detect micro-expressions in the first place, this descriptor capturing changes in motion was perfect.

Regarding other features, audio features were collected as well through the Mel-frequency Cepstral coefficients which is a tool widely used for automatic speech recognition. Instead, transcript features were collected through the Global Vector For Word Representation (Pennington, Socher & Manning, 2014): it encoded the entire set of words featured in the video scripts to a single fixed-length vector. The last step consisted in aggregating a variable number of features inside another fixed-length vector: the number of features was different for each video, so a Fisher Vector was used to accomplish the aggregation task.

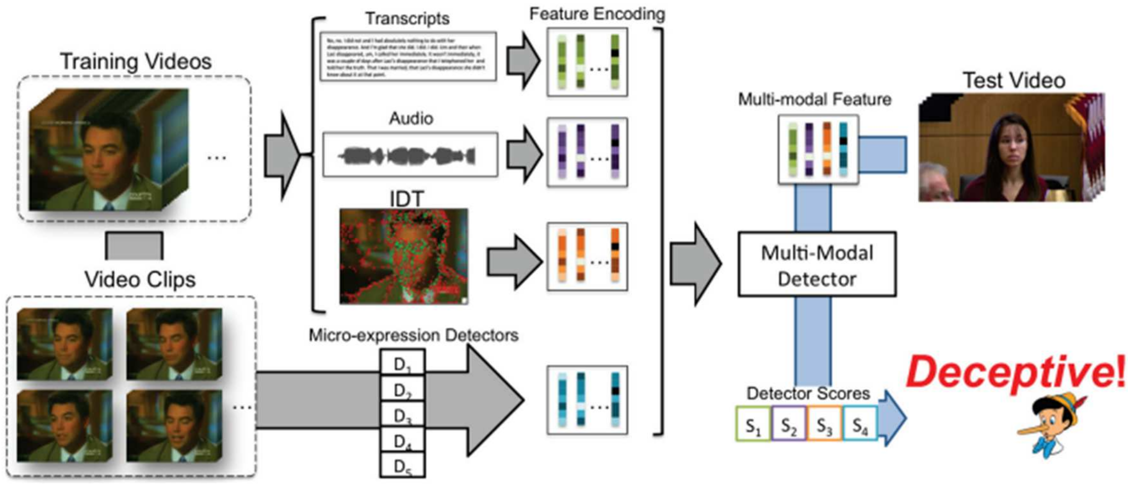


Figure 2.9. *the automated deception detection framework is depicted (Wu, Singh, Davis & Subrahmanian, 2018).*

Low-level visual features were used to train the micro-expressions detectors, and then the predicted scores of this performance were translated into high-level features for predicting deception: all the video clips of the database were divided into short fixed-duration video clips that were manually annotated with micro-expressions labels.

Across all the modalities, the highest “Area Under the Curve” (AUC, 0.8773) was reached after late fusion, which employed all the modality features and a linear classifier SVM. If micro-expressions were taken as a singular type of feature the AUC remained satisfactory but fell to 0.7502. Not confirming previous results, human performance on this recognition task showed an AUC of 0.8102, and so above the chance level. Researchers justified this late evidence by looking at the dataset: they defined it as relatively easier than previous studies where the prediction of human beings was almost chance (Wu, Singh, Davis & Subrahmanian, 2018).

The work from Wu and colleagues is not the only one to have employed the ground-breaking dataset of Pérez-Rosas et al. In this regard, studies that recently have disrupted this database are discussed in the next section.

Jaiswal and colleagues (2016) used the same 121 videos and analysed them through OpenFace software: it allows the extraction of facial features by even considering those transitions between them that are observable in video segments. From all the transcripts, linguistic and acoustic features were extracted too, and eventually they were combined with visual ones by using the early feature approach (or feature-level fusion approach). An SVM classifier was used, reaching an accuracy of around 78% over full videos; on the contrary, human judgment accuracy never went above 60%.

On the same data set, Krishnamurthy and colleagues (2018) used a multi-modal feature extraction approach in which visual features were notably extracted through a 3D convolutional neural network (3D-CNN). To date, a 3D-CNN system achieves state-of-the-art results in the classification field of tridimensional data, thanks to the combination of both image features and spatiotemporal features. Furthermore, a multi-layer perceptron (MLP) with hidden layers followed by a linear output layer was used. The combination of video, audio, text, and prominently micro-expression features lead the MLP to obtain an accuracy of 96.14% (ROC/AUC: 0.9799). The AUC when micro-expressions were used as a singular feature reached the 0.7512 value (Krishanamurthy, Majumder, Poris & Cambria, 2018).

Avola and colleagues (2019) used the OpenFace framework to extract AUs from the same dataset (2015) while an SVM was trained with a radial basis function kernel to extract and categorize AU (truthful or deceptive statements). They reached an accuracy of 76.84% outperforming (by a few units) those obtained by previous works using different algorithms (i.e., Wu et al. reached an accuracy of 75% using a linear SVM, whereas Pérez-Rosas et al. reached a 76% accuracy with the Random Forest algorithm; Avola, Foresti, Cinque & Pannone, 2019).

Rill-Garcia, Jair Escalante, Villasenor-Pineda e Reyes-Meza (2019) fed videos from the same database to OpenFace 2.1.0. and collected features from acoustical and textual modalities too. In addition, a new and smaller database was created from Mexican people speaking in Spanish about sensitive topics and personal topics. Among the court dataset, fusion methods (i.e., exploiting a different set of extracted features) reached their peak at around 0.645 AUC, whereas on the Spanish dataset it was around 0.631. Surprisingly, in this latter dataset, the best-performing feature was the gaze direction (0.769). However, the authors explained this outcome by mentioning the small number of training instances. Within the same work, the authors conducted another experiment by exploiting a Long Short-Term Memory (LSTM) whit 200 hidden units. LSTM was used as a way to include the temporal sequence natures of videos and computing features at the frame level. This architecture was tested with visual and acoustical features only achieving respectively an AUC of 0.560 and 0.730 on the court database. For what concerns the Spanish video clips, disappointing results were obtained, sticking to 0.38 and 0.294 values (Rill-Garcia, Jair Escalante, Villasenor-Pineda e Reyes-Meza, 2019).

Ding, Zhao, Lu, Xiang & Wen (2019) focused their work on a novel face-focused cross-stream network, the FFCSN. They analysed both facial expressions and body movements and trained an R-CNN for classifying the areas from each frame into objects (i.e., regions of interest) and refining the boundaries of these regions. Through this deep learning procedure, Ding and colleagues got an accuracy of 93.16% just considering the visual modality. When looking at the model using visual, acoustic, and verbal modalities the accuracy raises to 97% (Ding, Zhao, Lu, Xiang & Wen, 2019).

Finally, Monaro, Maldera, Scarpazza, Sartori and Navarin (2021) proposed a work (resembling a precursor of the current research) in which different extraction methods (i.e., improved dense trajectories and OpenFace) and machine learning techniques (support vector machine and deep neural networks) were applied to the analysis of facial micro-expressions for the identification of liars. Results from human naïve judges were collected too and compared with those of AI procedures. Notably, the database employed (previously collected by Monaro et al., 2020) was based on interviews in which a technique to increase liars' cognitive load was implemented: (Hartwig and Bond, 2007; Monaro et al., 2020a). Two conditions were formed: in the "truth-teller" condition participants were asked to recollect memories from a holiday taken place in the last 12-18 months, whereas in the "liar" condition they were asked to report an imagery holiday. Moreover, in each interview and condition, unexpected questions were asked (alongside free speech) to increase the cognitive load of the participants. The final data set of videos contained free speech and responses to unexpected questions. Although the stakes were low, creating a fictional holiday memory was assumed to involve the same cognitive processes as those observed in the setting of a criminal investigation when a false alibi is fabricated. For what concern the employed machine learning methods, three different approaches were followed: a linear SVM classifier was fed through both common algorithmic features extraction techniques and higher-level features (extracted through OpenFace software). The third approach used OpenFace as well, but there the SVM was replaced with a more complex LSTM network classifier. A fully neural approach was considered too through the mean of a 3D convolutional neural network (C3D): this network is directly fed with raw and unlabelled data (videos of interviews) and learns automatically during training the features of the input.

Checking the results, the SVM classifier obtained better outcomes when features were extracted with OpenFace (AUC of 0.78 per the "unexpected question" data set and of 0.72 for the free speech condition). Instead, the LSTM network paired with OpenFace features obtained a maximum AUC of 0.72 when paired with unexpected questions. Finally, testing the data sets of free speech and settled questions with the C3D network achieved respectively the AUCs of 0.64 and 0.75. Accordingly, the performance of the classifiers was better when dealing with unexpected questions: when presented with unexpected questions liars experienced an increased cognitive load, leading them to show more deception cues. As explained by previous studies, human judges didn't diverge from the chance level when attempting at catching lies, performing worse than how machine learning methods do (Monaro, Maldera, Scarpazza, Sartori & Navarin, 2021).

3. AIM OF THE STUDY AND METHODS

3.1. Aims of the study

The current study is developed following the model and evidence collected by Monaro and colleagues (2021), with the scope of highlighting if micro-expressions analysis can be beneficial for detecting lies. In particular, the interest is focused on observing and evaluating both human judges and machine learning methods' performances when attempting to differentiate between honest subjects and liars. Their performances are compared in terms of accuracy (or the fraction between the number of correct detection guesses and the total number of guesses). We expect that machine learning (i.e., deep artificial neural networks) outperforms human judges' lie detection scores, in accordance with previous studies (e.g., Bartlett et al., 2014; Pérez-Rosas, Aboulenien, Mihalcea & Burzo, 2015; Wu et al., 2018; Monaro, Maldera, Scarpazza, Sartori & Navarin, 2021). It is expected that human beings' lie detection performances do not statistically and significantly diverge from chance level (i.e., almost 50% of guesses, DePaulo et al., 2003; Bond & DePaulo, 2006; Porter & Ten Brinke, 2008; Bartlett et al., 2014; Pérez-Rosas et al., 2015; Curci et al., 2019).

Besides comparing the lie detection performances of humans and artificial neural networks, the experimental procedure is even constructed for understanding if the type of lie (i.e., resembling the association respectively to a bigger or a smaller cognitive load) can influence its detection. Indeed, participants are presented with images having different levels of complexity: it is hypothesized that the complexity of each image influences the ease with which a false statement is constructed and in turn, the ease with which it is detected (or, correctly labelled as a truth or a lie). The hypothesis is that describing a complex figure should be more cognitive demanding compared to simple figures, increasing the cognitive load experienced by the liar while deteriorating his/her ability to hide what is concealed. Both humans and neural networks are assumed to take advantage of more complex figures while playing the role of judges; facial leaks and especially micro-expressions may be more evident and more likely to come out, simplifying the lie detection procedure. To the best of our knowledge, this study is the first one to consider the influence of the type of lie on its detection through facial micro-expressions. To date, no studies have considered how the complexity of the target of the lie may influence the cognitive load experienced and may facilitate deception detection in both human judges and machine learning methods.

Briefly, regarding the two presented objectives, it is expected to achieve results along the lines of those of previous lie detection research, both for what concerns human judges and automatic learning techniques. Furthermore, new evidence regarding the importance of strictly considering the characteristic of deceptive statements (complex vs. simple) is expected to be highlighted. Broader implications of the study are identifiable in the plausible effects for the forensic fields, that is the development of an artificial tool meant to ameliorate lie detection in legal contexts.

3.2. Method and dataset

The research is promoted by the "Human Inspired Technology Center" (HIT) of Padua University and run by professional figures from both the departments of general psychology and Mathematics "Tullio Levi-Civita".

The full process concerning data acquisition was carried out through the employment of an ad hoc developed smartphone application (i.e., "Lie detector" app); 110 completed matches and 330 total videos were collected, having an average length of 8.82 seconds (*DS=4.1*) and all resembling the recording of the experimental participant's face intent on verbally describing a given image. Images were presented via the screen of a smartphone provided by the research team, therefore participants just had to look at it and freely produce a description on it. Crucially, participants were left free to decide if producing a truth or a deceptive statement, that is respectively to produce an authentic and consistent description or a fictional one. Enough to a description for being categorized as "false" was the alteration of a single small visual detail. A database of images was upstream uploaded on the backend of the application, containing a total of 150 images. Images were divided into two classes of 75 stimuli each: simple and complex. Simple images always feature a unique subject (everyday objects, animals, or food indistinctly) placed on an empty white background; on the contrary, complex images feature detailed landscapes and multiple subjects. An example of a simple and complex image is reported in figure 3.1.

The categorization led to a final dataset characterized by 172 false descriptions and 167 true descriptions. Each video was then submitted to the "judgment" of another participant who was asked to categorize the veracity of the description (i.e., true or false) accordingly to his/her own perception. Additionally, the same videos have been used as training- and test-dataset for the learning process of artificial neural networks.


Figure 3.1. *Examples from the database of images used as experimental stimuli are reported, respectively a complex one on the left side and a simple one on the right side. Even though they share the central theme, the compounding features are widely different. Within the "complex figures" categorization images characterized by multiple subjects and details on the structured background are allowed. On the contrary, as depicted by the image on the right, "simple images" are characterized by a unique subject (both objects and animals) on a strictly white background (University of Padua, department of computer sciences/general psychology, 2022).*

3.2.1. Participants

103 individuals took part in the experimental procedure, although 8 of them had to be discarded due to their non-compliance with the experimental instructions. In total, 95 participants were ascribed to the final experimental participants' group, 38 of them were females while 57 were males, aged between 19 and 32 ($M = 22.9$; $SD = 2.74$). Their years of education vary between 13 and 18 ($M=14.98$; $SD=1.7$): 38.94% have a high-school diploma, 52.63% have a bachelor's degree and the remaining 8.42% have a master's degree. The discrepancy between participants' total number and videos is explained by the possibility for a player to take part in more than a single game.

3.2.2. The "lie detector" application

The "lie detector" application was built on the work of two students (Dametto Alex & Poloni Mattia) from the "Computer Science" master's degree in Padua. Professors Navarin, Gaggi, Palazzi and Monaro took part as supervisors. The development of the first version of the application started in February 2020 and after several changes, a final version was completed in March 2022. Finally, the "lie detector" application was programmed in the form of a 1vs1 game (Figure 3.1) where two competing players attempt to overcome each other through lie detection guesses. The final version features an admin dashboard that makes it possible to check game statistics (e.g., the total number of players or games played) and to remotely manage the data collection: modifying and cancelling specific data was made possible, as well as checking users' info and reports or alerts concerning the flow of their matches.

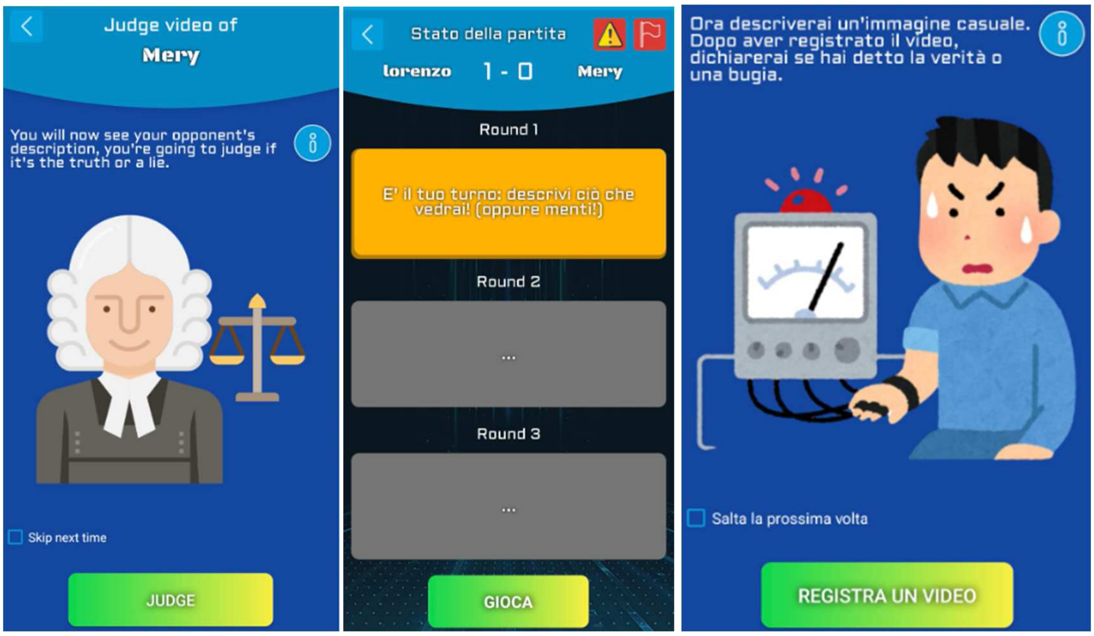


Figure 3.1. *Three dashboards from the "lie detector" application (A, B, C) are reported. The game flow is set around a match between two participants who alternate roles as "judge" (A) and "speaker" (C). The speaker can either produce a false or a true statement while the judge-role player has to guess whether his/her opponent is lying. Each match of the game consists of three rounds, meaning that each player will act as judge and speaker three times. A dashboard depicting the resume of the current match (B) is always available while playing (University of Padua, department of computer sciences/general psychology, 2022).*

The application is meant to run on Android devices, and it was developed in Java by using the developer tools "Android Studio"; otherwise, data were saved and stored through the database programme "MongoDB" that collects all the information related to users, matches, reports and videos. The back-end portion of the software that acts as an interface between the android application and the database was developed in "JavaScript" through the framework "NodeJS". Finally, the admin dashboard was developed as a final step procedure in "React" (i.e., a front-end library from JavaScript that is commonly used for building user interfaces, Figure 3.2).

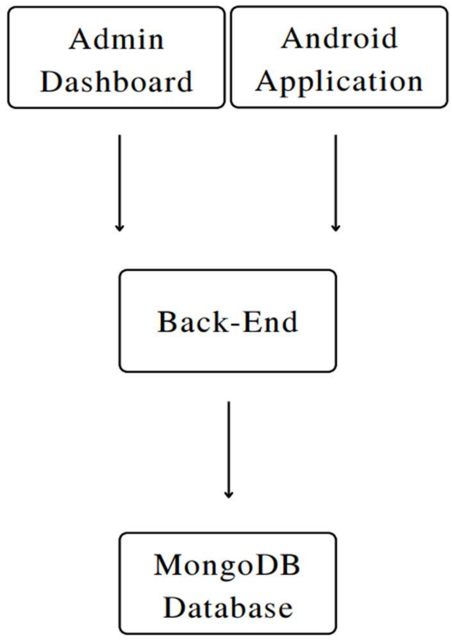


Figure 3.2 *The technology stack that was used to run and built the "lie detector" application (University of Padua, department of computer sciences, 2022).*

The latest works on the application were responsible for preventing it from exclusive use, beside developers the "lie detector" app was indeed intended for all external users. Thus, a server from the University of Padua was chosen for the back-end, allowing to gather a wider amount of data. Eventually, an APK was developed in an effort to make it possible to install the application on potentially every user's phone. As a final note, all the versions of the application were programmed following the guidelines regarding usability, the graphical interface, and the data processing methods as well.

3.2.3. The data acquisition and the experimental procedure

The study in question was revised and approved by the ethics committee of the HIT centre on the 11th of march 2022. Participants were collected between the 29th of April 2022 and the 20th of June. They were all volunteers who were directly asked whether they want to participate on-site. As a very first step, each participant read and hand-signed an informed consent in which he/she authorizes the processing of his/her data and agrees to the confidentiality treat of the same. Two smartphones provided by the University and the research team (with the lie detector app pre-installed) were used for the data acquisition, thereby throughout the gameplay, both participants were able of playing by interacting with just one single device each. After picking up the smartphone, the participant received the experimental instructions as follows.

Firstly, each subject was asked to create an in-game account within the application (Figure 3.3A) that consists in uploading an e-mail, a username and a password, and personal information as well (i.e., age, sex and scholarship). The accounts are unique and were intended for the correct uploading and collection of the basilar information needed for statistical purposes. As soon as the two subjects sign in with their account it was possible to “create” a match by an online paring of their accounts. At this point, participants were told that the game consists in describing an image shown on the screen of their smartphone; moreover, they were informed as well about the possibility of choosing if describe it truthfully or by lying.

After a countdown of 5 seconds, a recording of the participant’s face describing the image starts: 15 seconds is the maximum length of the recordings that as soon as it is interrupted can be sent and saved to the database furnished by the University. However, once the video is recorded the participant has the will to whether send, re-shoot the video or leave the gameplay and avoid any data storage. As a very final step, the participants are asked to genuinely categorize their description by specifying through a “Truth”/” Lie” button-alternative whether the statement is true.

Subsequently, the second player received and reviewed the recorded video of his/her opponent on the device: he/she is instructed to guess whether the description is a lie and to confirm the decision through a “Truth”/” Lie” button (Figure 3.3B). No instructions are given concerning which strategies the player should use to make the guess more accurate. Then, the gameplay continued for three total rounds during which the two players alternate themselves in the role of “detective” and “speaker”. As anticipated, all videos were saved on the university server and later on were used for training and augmenting the performance of a similar AI stack as the one previously developed for the work of Monaro and colleagues (2021). Participants’ lie detection guesses were stored as well and employed to gather statistical information concerning human judges’ performance.

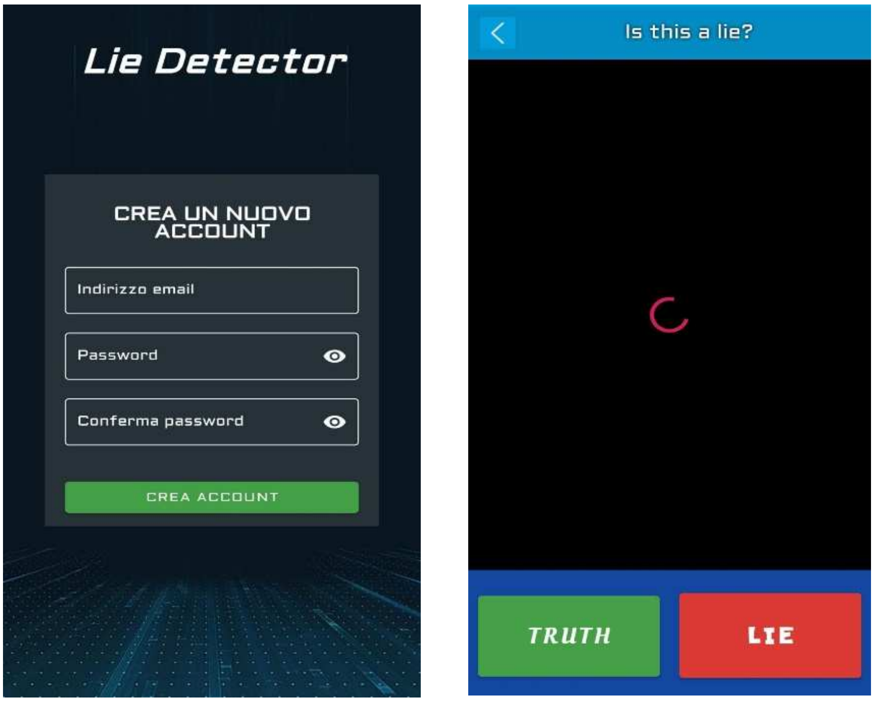


Figure 3.3. *(A), on the left, reports the in-game dashboard for the creation of the account: it was used as a unique identifier for the subject/player during data processing. (B), on the right, reports the frontend two-button-fashion for the detective's categorization. Usually, within the depicted black space, the video description of the other player is reported allowing the detective to start the in-game lie detection procedure (University of Padua, department of computer sciences/general psychology, 2022).*

3.3. Machine learning methods

The entire database of videos collected on participants' verbal descriptions has been used to train three different automatic learning algorithms (machine learning). Generally speaking, machine learning can be defined as the field of study that enables computers to learn without being explicitly programmed, while relying on different algorithms to solve data problems. Within a complex task such as the analysis of micro-expressions, a tool capable of handling the data more efficiently is in demand (Mahesh, 2020). Thus, the idea behind it is to use machine learning models for determining and analysing the facial movements (and micro-expression) of a person while he/she is potentially induced to lie: by comparing the newly obtained information with previously stored and collected data concerning individual subjects' visual morphology, the computer is eventually expected to conclude whether a person is lying (Azhan, Zaman & Bhuiyan, 2018).

The workflow that leads to automatic learning can be expressed under two opposite tasks of learning, respectively supervised learning and unsupervised learning. Here, supervised learning was employed, since the current automatic learning was based on constructing a function that maps an input to output following the presentation of input-output pairs (Figure 3.4). These types of learning algorithms required external assistance and labelled training data resembling a set of training examples. As regards the input-body of information, in supervised learning the labelled dataset is divided into training and test dataset. The training-set and correlated steps of learning are designed to let the algorithms learn some kind of patterns, meanwhile, the test-set is rather intended for checking the prediction or classification performances that the model is expected to show at the end of the learning process (Mahesh, 2018).

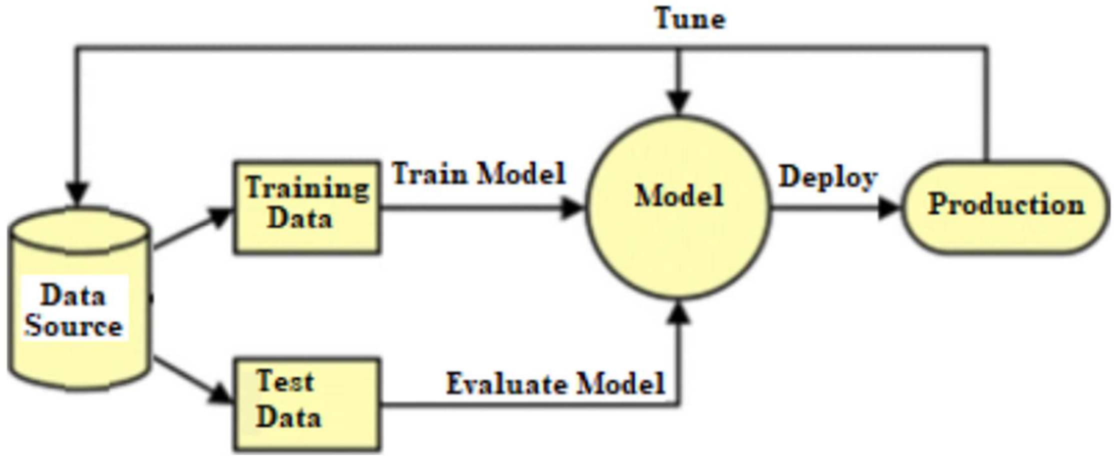


Figure 3.4. *The supervised learning stack is reported. The learning process is designed to achieve a tune between the data given in input (“data source) and the final answer produced by the model (“production). Training data and test data are both fundamental for the correct flow of the learning process, respectively they let the model “understand” the desired patterns and evaluate whether it succeeds in carrying out the desired task (Mahesh, 2018).*

Worth a mention is the way the training dataset and the test dataset were guaranteed to differ: each participant’s face appeared at least three times (as the number of rounds of a single match) in the whole dataset while there was no maximum, considering that there was no official limit in the number of playable games (even though a maximum of three matches was suggested and observed on most occasions). Participants playing more than one game were instructed to log in always with the same account, the uploaded e-mail served as an identifier of the participants, making it possible to have a training set and a test set made up of videos of diverse faces.

Crucially, features other than micro-expressions could have been considered and used as sources of information for the machine learning analysis, given that the act of lying is a multimodal process per se that embraces different physiological responses. However, in this study, we decided to focalize the analysis just on facial micro-expressions, with the purpose of considering further features in future works. Nevertheless, the already discussed studies (e.g., Pérez-Rosas, Aboulenien, Mihalcea & Burzo, 2015; Monaro, Maldera, Scarpazza, Sartori & Navarin, 2021) have highlighted how fundamental features of micro-expressions can be for machine learning analysis, even when considered individually successful results are indeed obtained. Therefore, during the pre-processing steps (meant for adapting the dataset to the selected algorithms) facial behaviour was exclusively considered and analysed. Video processing has been carried out using "OpenFace" software (Baltrusaitis, Mahmoud & Robinson, 2015; Baltrusaitis, Zadeh, Lim & Morency, 2018). Through it, the speaker's face was highlighted ("cropped") from the surroundings and 17 AUs were extracted: the derived AUs from every video-frames were then measured on a scale from 0 to 5, each resembling one of the anatomical units that underlie facial movements.

3.3.1. An overview of the employed automated models

As a baseline and starting model, an SVM classifier was employed. The SVM classifier, first proposed by Cortes and Vapnik (1995, Figure 3.5), was originally understood as a learning machine for two-group classification problems; it was a supervised learning model in which learning algorithms analyse those data used for classification and regression analysis. However, besides that, nowadays an SVM model can efficiently carry out even non-linear analyses by disrupting the so-called kernel trick (i.e., implicitly mapping the inputs into high-dimensional feature spaces). The logic behind it is "to draw" margins between the classes of stimuli in a way that the distance between classes and margin is maximum so that the classification error is minimized (Mahesh, 2018).

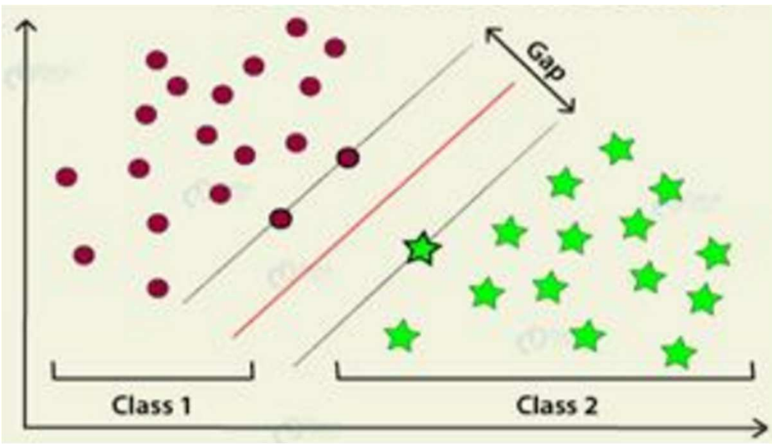


Figure 3.5. *A graphical representation of a support vector machine classifier (Mahesh, 2018).*

A subsequent step within the machine learning procedures stack was the implementation of a "Transformer" model (Vaswani et al., 2017). Originally, the transformer was developed as a simplification of earlier up-to-date neural network approaches for sequence modelling and transduction problems (e.g., recurrent neural networks, long short-term memory and gated recurrent neural networks), and was designed to solely rely on the attention mechanism. In particular, it was proposed as a solution to the sequential nature of the cited recurrent models: thus far, the "time" variable was computed starting from sequences of hidden states (h_t) and by computing a function of the previous hidden state (h_{t-1}) alongside the input for position t. Vaswani and colleagues (2017) assumed that this sequential nature precluded a complete and necessary parallelization of the training examples, which is a critical issue in presence of stimuli with a long sequence length. Through the Transformer, a global dependency between input and output is drawn by relying entirely on attention and dispensing with a recurrent architecture. This simplification is achieved by three main components: an encoding component, a decoding component and a connection between them (Figure 3.6; Vaswani et al., 2017).

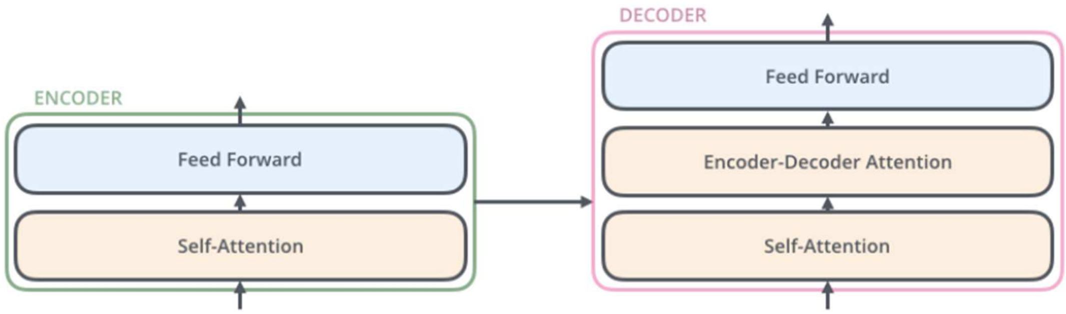


Figure 3.6. *A simplified representation of the three main components of the Transformer is depicted. Each encoder component is divided into two sub-layers: at first, the input of the encoder flow through the self-attention layer that is responsible for evaluating and considering all the sub-part of the input (other than the precise stimuli per se). The output of this sub-layer is then fed to a feed-forward layer that communicates with other encoders. The decoder component does have the same layer as the encoder, but in addition, it involves an attention layer. Thanks to this item, the decoder focuses on a relevant part coming from the output of the encoder stack (Vaswani et al., 2017; Adapted from Alammar, 2018).*

The third and final effort was the implementation of a "Siamese Network" architecture with a triplet loss function. Siamese Networks, differently from most cutting-edge neural networks, perform well even in presence of a few amounts of data; put differently, by exploiting a Siamese network architecture it is virtually possible to leverage a limited amount of labelled stimuli (here, videos) to effectively detect lies (Jiabo, 2021). This type of architecture contains two (or more) equal sub-networks that show an identical configuration of parameters and weights. The updating of the parameters occurs in parallel and in a specular way in both sub-networks, making it possible to look for similarities across two different stimuli (fed as input to one of the two sub-network respectively).

In this specific case, the Siamese network was trained under the “triplet loss” function. Firstly, to construct a triplet, a training example is randomly set as the anchor, while the other two examples are selected as the “positive” and the “negative”: the positive shares the same label as the anchor, whereas the negative, does not. The loss function optimizes the embedding model weight such that vectors (i.e., videos) of the same truth-lie category (the anchor and the positive) are represented closer while vectors that are different are further away in the embedding space. When the learning process is completed, an average of all the representations of truth-lie video is computed separately; to categorize a new video it will be chosen the closest spatial representation category (Jiabo, 2021).

4. DATA ANALYSIS AND RESULTS

4.1. Lying behaviour

Out of 330 recorded videos, the frequencies of true statements and false statements resulted balanced: participants lied on 167 occasions while preferring to tell the truth in 163 of them (respectively, 50.6% and 49.4%). Thus, on average, referring to the single games, participants resolved to lie on 1.52 rounds out of 3 while telling the truth on 1.48 rounds (sd= ± 0.82).

Moving forward, another point of interest was understanding the type of "lie behaviour" the participants adopted, whether they were presented with a complex or a simple image (Table 1). A discrepancy in the disposition to tell lies accordingly to the type of image was observed since participants lied 55% of the time when presented with simple images and 46% of the time when presented with complex images. However, the percentage difference resulted to be statistically non-significant: a chi-squared test was computed to evaluate the relationship between the type of image and lie behaviour, $X^2_{(1, N=330)} = 2.0206$, $p = .1552$.

	Videos categorization		
Image Complexity	*False*	*True*	*Total*
Complex	72	84	156
Simple	95	79	174
Total	167	163	330

Table 1. *The contingency table concerning the complexity of the image (complex/simple) and the type of statement (true/false) is reported.*

To deepen the influence of image complexity, video's lengths (in seconds) were collected and compared too. Briefly, an ANOVA was calculated and the effect of the type of statements (truth/lie) and image (complex/simple) on videos' duration was acknowledged (Table 2). There was a significant main effect for image complexity ($F_{(1, 326)} = 8.854$, $p = .003$) suggesting that the description correlated to complex images lasted more than those correlated to a simple image.

Cases	Sum of Squares	df	Mean Square	F	p	η^2
Image Complexity	144.707	1	144.707	8.854	0.003	0.026
Videos: Truth/Lie	19.848	1	19.848	1.214	0.271	0.004
*Image Complexity * videos Truth/Lie*	0.183	1	0.183	0.011	0.916	3.324e - 5
Residuals	5328.214	326	16.344			

Table 2. *The ANOVA regarding the effect of the image complexity (complex/simple), the video categorization (truth/lie) and their interaction on the duration of videos is reported.*

No statistical significance was observed for both the effect of the video categorization taken singularly and the interaction between it and image complexity (respectively, (F (1, 326) = 1.214, p = .271); (F (1, 326) = 0.011, p = .916). The post hoc analysis is reported in Table 3.

		Mean Difference	SE	t	p_{Tukey}
Complex, False	*Simple, False*	1.378	0.632	2.182	0.130
	Compex, True	-0.446	0.649	-0.686	0.902
	Simple, True	0.838	0.659	1.272	0.581
Simple, False	*Complex, True*	-1.824	0.605	-3.013	0.015
	Simple, True	-0.540	0.616	-0.878	0.816
Complex, True	*Simple, True*	1.284	0.634	2.026	0.180

Table 3. *A multi-comparison was conducted throughout Tukey's range test concerning the interaction effect between Image Complexity and videos' categorization.*

4.2. Humans' lie detection performance

The analysis of human judges' performances was then centred on the accuracy of their guesses. Overall, participants succeed in categorising correctly the video descriptions on 57.3% of the occasion, showing no difference between the number of guesses made on truth statements and those made on lies: the percentage of true negatives (i.e., a true statement categorize as such) was 57.1%, while the percentage of true positives (i.e., false statements categorize as such) was 57.49%. In other words, participants' detecting performances resulted in not being influenced by the type of statement (truth/lie) they were asked to evaluate. Finally, by considering each round cumulatively, a little amelioration of the detection performance was appreciated (Round 1:54.9%, Round 2: 56.2%, Round 3: 58.4%; Figure 4.1). Briefly, in all circumstances, participants' scores did not diverge significantly from chance level.

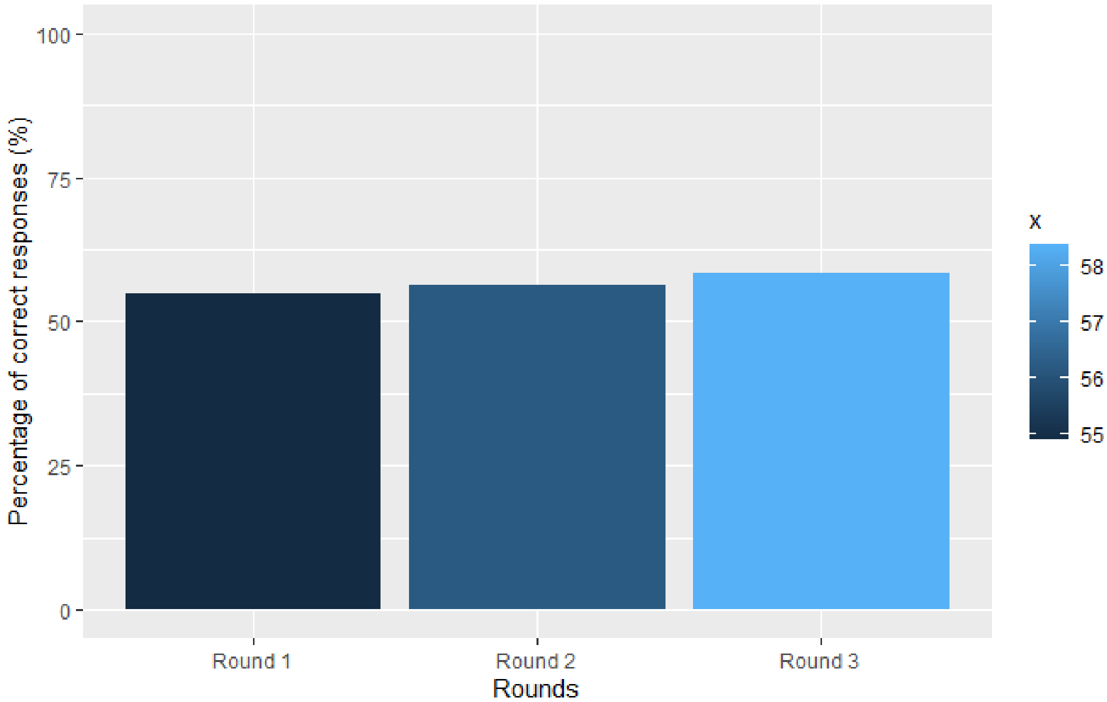


Figure 4.1. *The graph reports the detection scores in percentage and across all the three rounds taken cumulatively. Crucially, these results were calculated by excluding all those matches following the first one played by each participant; exclusively by doing so it was possible to appreciate the trend of the performance within single games. Eventually, an amelioration of the performance is observed, suggesting the possibility that getting used to the task and the in-game opponents have beneficial effects on the performance (making the task easier to complete).*

An analysis of the influence of image complexity on human judges' accuracy was conducted as well. A chi-square test of independence was calculated to examine the relation between the variable- image complexity and the categorical variable resembling humans' accuracy. The outcomes did not suggest a statistically significant relationship, $X^2_{(1, N=330)} = 0.006$, $p = .939$.

	Humans' accuracy		
Image Complexity	*Right*	*Wrong*	*Total*
Complex	89	67	156
Simple	100	74	174
Total	189	141	330

Table 4. *Human judges' accuracy scores are reported concerning the type of image they are presented with.*

4.3. Machine learning results

The dataset fed to the three machine learning methods followed the addition of 9 extra videos that previously were discarded from the statistical analyses concerning human judges' performances. In total 339 videos were employed and divided across both training set and test set: within AI analyses it was preferred to use even those videos (excluded by humans' analyses in order to examine lying behaviours under a complete picture) that did not belong to completed game-matches, that is videos belonging to matches in which at least 1 out 3 round were discarded for a non-compliance to experimental rules.

Videos from the dataset were divided into 10 sub-groups to allow 10-fold cross-validation: 5 groups consisted of 10 participants while the remaining five of 9 (Table 5).

Groups	Truths	Lies	Total
0	21	17	38
1	16	14	30
2	18	18	36
3	20	15	35
4	13	20	33
5	19	14	33
6	13	16	29
7	23	23	46
8	13	19	32
9	11	16	27

Table 5. *A balanced breakdown of the participants has been sought to achieve. The table reports in the second and third column respectively the number of truth and lie videos belonging to each fold and the overall in the last one.*

As anticipated, the first employed model was an SVM classifier where first and second-grade statistics of the AUs were implemented as features of the model. Each video was accounted for as a 34-sized vector, the first 17 dimensions were provided by the mean on all the frames of the single AUs, while the other 17 by their variance. Hence, the cross-validation was performed by using 9 groups as the training set and the remaining ones to test the performances of the model; the whole procedure was repeated for evaluating each group. Mean accuracy of 0.5750 was achieved (SD = 0.0502).

For the Transformer model, the procedure was similar: cyclically the i-th group was used as the test set, the i+1-th as the validation set and the remaining 8 as the training set. The hyper-parameters were fixed as equal in every single fold to shorten the computing time. A 0.5640 mean accuracy was achieved (SD= 0.0872).

Finally, the triplet loss function through which the Siamese network was trained consisted in feeding the last model with three examples at a time: videos in the same category were represented in the loss function space close to each other and distant from the latter belonging to the other category. Mean accuracy of 0.5572 was achieved (SD = 0.0600). In Table 6 a summary of the results is reported.

Test Fold	SVM	Transformer	Siamese
0	0.5263	0.5263	0.5789
1	0.6000	0.6000	0.6000
2	0.5000	0.5833	0.4722
3	0.6000	0.6000	0.6857
4	0.6364	0.3636	0.5758
5	0.5758	0.6970	0.5152
6	0.5517	0.6207	0.5517
7	0.6304	0.6304	0.5000
8	0.5000	0.5000	0.5000
9	0.6296	0.5185	0.5926
Mean	0.5750	0.5640	0.5572

Table 6. *The accuracy (0-1) of the performance of the three machine learning models employed is reported on a scale between 0 and 1. The Data are reported following the folds, with a latter row dedicated to the average value of the accuracy.*

5. DISCUSSION AND CONCLUSIONS

5.1. Evaluating human judges' performances

The task carried out by the participants while acting as "in-game detective" was nothing more than a lie detection procedure and as such, their performances did not diverge significantly from those observed in previous studies. Here, human judges showed an accuracy (57.27%) that coherently is close to the chance level. These results are like the ones published by Monaro and colleagues (57%; 2021) while proving to be just slightly higher than Pérez-Rosas et al. (51%; 2015) and Bartlett et al.'s findings (51.9%; 2014). Notably, it is appropriate to consider that in the current experimental procedure, participants that took part jointly in the 1vs1 game/task occasionally knew each other: potentially, performances might be influenced by this aspect, assuming that people knowing each other should be facilitated in understanding both the verbal and non-verbal behaviour of their opponents.

Beyond validating that human scores do not diverge from chance level when attempting to spot lies, another source of interest was figuring out which might be the influence of the type of lie on its detectability. Here, the type of lie was assessed in terms of the associated cognitive load; put differently it was assumed that lies correlated to complex images were more cognitively demanding and in turn more prone to be detected. Ekman and his colleagues coined the term "leakage" to refer to a liar's unintentional betrayal of truth that emerges as a result of high cognitive load (and emotional load too; Clancy, 2009). Up to now and to our knowledge, this was the first attempt to establish whether different types of lies were associated with different detectability levels. However, comparing the detection accuracy correlated to complex and simple images' descriptions no statistically relevant difference was highlighted, since the accuracy associated with simple images was 57.1% and the accuracy associated with the complex image was 57.5%.

Despite being out of our expectations, these findings might be explained because of the few constraints participants were asked to adhere to while producing their descriptions. Subjects had indeed no restrictions concerning their verbal description, meaning that when presented with a complex image rather than a simple one, the sought cognitive load might be easily averted. Lies on complex images were most of the time built with apparent ease by focusing on few details or creating new yet simple ones: to simplify the task, fictional descriptions were most of the time complete fabrications (having nothing in common with the stimulus they are presented with) or partial truths (that is, altering just one tiny detail out of several to obtain what researchers might address as a "stretched truth"; Bryant, 2008). Handling all the several details of a complex image finishes to be unnecessary in such cases (describing an image truthfully does not require any manipulation process), and causes the subject to not experience an excessive cognitive load.

Finally, the difference in length between videos associated with complex images and simple image descriptions is worth a mention. A statistically significant difference was appreciated, meaning that the description of complex images lasted more. Different explanations can be drawn to explain this result, starting from the obvious presence of much more information to be described compared to simple images. Moreover, considering plausible the onset of an actual cognitive load, a greater latency to begin describing complex images should be acknowledged. If this is the case and a pragmatic difference between complex and simple description exists, it would be reasonable to address an explanation that goes towards the already discussed inability of human judges to identify and disrupt the few and available cues to deception (Hartwig & Bond, 2011).

5.2. Machine learning methods results and human judges: an explanation of the collected evidence

The consistency with previous studies observed in humans scores did not apply to the results collected through machine learning procedures (Owajyan, Kashour, Haddad, Fadel & Souki, 2012; Su & Levine, 2014; Bartlett et al., 2014; Pérez-Rosas, Aboulenien, Mihalcea & Burzo, 2015): machine learning techniques were indeed expected to outperform human judges' scores, with a special interest in reproducing the results obtained in the precursor study by Monaro et al. (2021). Disappointingly, here the performances of machine learning models stand around the chance level (Table 7), posing the issue of why a similar AI stack as in Monaro et al. (SVM classifier, LSTM and C3D) obtained better results (showing overall accuracy around 0.70).

Models	Accuracy
SVM	0.5750
Transformer	0.5640
Siamese network	0.5572
Human Judges	0.5727

Table 7. *A comparison between human judge's performance and machine learning methods accuracies (in a range from 0 to 1) is reported.*

The motives behind the disparity in these performances can be addressed in light of the unconventional experimental paradigm and setting here employed. Firstly, data were collected within uncontrolled environments; namely, participants took part in the experiment directly on sight and had to deal with external variables while being presented with the task. A certain amount of background noise was given for guaranteed since the development of the experimental design, although it was preferred to dispense with the laboratory to prevent as much as possible low levels of ecological validity. In addition, the idea of employing a game-like paradigm for collecting data was intended to reproduce the cognitive processes involved in the real act of lying and augment the level of experienced pressure. Nevertheless, the background noise might have ended up overly influencing the signal needed for the AI networks to perform: for instance, some of the collected facial expressions (and specifically micro-expressions) might have been elicited in response to an external stimulation rather than the effort to deceit.

A no minor source of concern comes from the design of the videos. Differently from previous research, in the current database videos were composed of few frames (with an average duration of 8.8s compared to the average duration of 267s of the stimuli from the study by Monaro et al., 2021). Thus, beyond a wider amount of videos collected, the amount of information fed to the AIs was arguably insufficient, given that the maximum allowed length of each was 15s: next to all the advantages they brought, deep learning architectures like the ones here disrupted, require a large amount of data, posing one of the most urgent issues within the domain of automatic deception detection, in other words, the availability of ad-hoc databases (Pérez-Rosas, Aboulenien, Mihalcea & Burzo, 2015; Hasan et al., 2019).

Lastly, it is worth mentioning that the experimental paradigm per se has allowed achieving some important evidence for future research lines, albeit it has not paid off in terms of results. In all the already mentioned studies concerning lie detection and machine learning methods, the deceits were shaped like lies concerning memory traces to be recalled. In Pérez-Rosas et al.'s study (2015) the database relied on court hearings where each involved subject knows exactly the event on which is expected to report. In Monaro and colleagues' study (2021), participants had to recall an actual memory trace and make a cognitive effort to insert fictional details. Thus, when asked to report their statements, they had to both recall the memory trace and modify it, in such a way that their cognitive load (and deceit leaks) might augment.

In the study at hand, there was no such effort related to memory recall. The participants lied directly at the time they saw the image, easing widely their cognitive effort and pressure, meaning that the likelihood for detection cues to emerge was reduced. Similarly, it might be the case that the pressure of performing as "good liars" and the associated arousal was lacking: here lying was a choice of free will (as in real case scenario, but crucially with no need of protecting own interests) made at the very last moment. Given the imposed short duration of the video, all the cognitive processes correlated to the act of lying might not emerge fast enough, rendering the signal including micro-expressions too narrow even for machine learning models. As a confirmation of this, occasionally during data collection it happened that participants resolved to lie only in the very last seconds of their recorded description by altering just a single detail of the image characteristics, rendering the on-set of lying-associated cognitive processes very subtle and the detectable signal for machine learning techniques too narrow to disrupt.

5.3. Significance of the current study and conclusion

The aim of this work was to compare the performances human judges and machine learning models (an SVM classifier, a Transformer and a Siamese Network) obtain when dealing with the same dataset. Regrettably, machine learning techniques showed poor performances when compared to those of previous studies, while not succeeding in outperforming human performances as expected. Moreover, this work represents the first attempt in the literature to ever try to deal with unstructured and spontaneous lies. The great body of work concerning automated lie detection processes has usually and appropriately focused its efforts on processing recordings of court hearings since the very last aim of lie detection research is providing a device capable of automatically detecting deception in professional settings. Among the exceptions, Monaro and colleagues (2021) worked on a new database based on pre-formed and structured reports concerning the description of a past holiday. Coherently to recreate a context close as much as possible to a court trial, these reports were assumed to involve the same cognitive processes as those observed when telling an alibi.

Having said this, involving datasets of non-planned lies might happen to be beneficial for lie detection advancements, especially as part of the generalizability of the results: within forensic contexts, lies are not always planned; deceptive behaviours may happen to emerge as a response to unexpected situations positing the demand for better performing AI models. Nevertheless, it has been highlighted how the experimental paradigm of the current study resulted in being overwhelming complicated for an AI to work correctly. To put this in perspective again, some serious limitations were presented. Conducting the experimental procedure outside of the laboratory surely influenced the control over external and undesired variables: participants' facial responses might be influenced by stimuli other than the presented image, questioning how strong might have been the influence of lying processes on their facial expressions. Alongside this, the employed machine learning models classically require a large amount of data to be effectively trained (e., Zhou, 2016); here, due to the short duration of the collected videos, this condition was not met, leading the available frames to be fed to the AIs to be arguably insufficient. Therefore, the evidence that should be considered while assessing the results presented is the following.

To date, machine learning models relying on micro-expressions analyses require to be fed with videos associated with "structured lies": to obtain acceptable levels of detectability, lies characterized by a wider temporal duration (i.e., more frames available as AI signal) and complete fabricated details (thus excluding partial or half-truth where an ambiguous level of truth may be found) are probably needed.

These findings are coherent even within Ekman's framework (e.g., Ekman, 2009): it has been already highlighted that micro-expressions if taken singularly, are not unambiguous indicators of deception. Then again, micro-expressions show neither a high temporal specificity nor stand as unequivocal deception cues, meaning that the onset of a micro-expression does not directly imply that deceptive behaviour is taking place. For this reason, too short videos or half-truths (i.e., white lies that are just partial truth) might even occur with no micro-expressions, leading the stimuli to be undetected by the employed neural network due to the lack of an actual signal to work on. In addition, complex and structured lies are usually more emotionally- and cognitively loaded (including memory recall processes), translating into a wider flow of leaks whose presence seems to be crucial for a straightforward performance of machine learning models. In this regard, in the experimental paradigm at hand, participants might have never felt under pressure: although the lie detection task was covert within a competitive 1vs1 game, the uncontrolled settings and the occasional familiarity between participants might have prevented task-related stress.

On the other hand, a source of interest for future lines of research might result be the experimental paradigm proposed in the form of a 1vs1 game. Creating such a competition between two participants might be an aid for emulating the kind of pressure or arousal experienced when telling a lie in a real case scenario (while keeping in mind the importance of avoiding recruiting subjects having familiarity with each other).

Then again, even if lie detection studies that exclusively consider the analyses of micro-expressions show promisingly results, it must be stressed that facial cues are just a portion of all the hints to deception we may elicit when lying: several “lie patterns” exist within our body and face under different shapes, suggesting the importance of striving to consider them all. We still do not know which pattern machine learning relies on when dealing with a lie detection task, but we do know that lying requires several and differing cognitive and emotional processes. However, among all the other cues, micro-expressions results are a pure predictor of a context (hot-spots) within which the subject is likely to be deceptive, and as such, it is still in demand to work on it.

REFERENCES

- A World of Lies. (2006). *Journal of Cross-Cultural Psychology*, *37*(1), 60–74.
- Abe, N. (2009). The neurobiology of deception: Evidence from neuroimaging and loss-of-function studies. *Current Opinion in Neurology*, *22*(6), 594–600.
- Abe, N. (2011). How the Brain Shapes Deception: An Integrated Review of the Literature. *The Neuroscientist*, *17*(5), 560–574.
- Abe, N., Suzuki, M., Mori, E., Itoh, M., & Fujii, T. (2007). Deceiving Others: Distinct Neural Responses of the Prefrontal Cortex and Amygdala in Simple Fabrication and Deception with Social Interactions. *Journal of Cognitive Neuroscience*, *19*(2), 287–295.
- Alliger, G. M., & Dwight, S. A. (2000). A Meta-Analytic Investigation of the Susceptibility of Integrity Tests to Faking and Coaching. *Educational and Psychological Measurement*, *60*(1), 59–72.
- Apfel, V. (2001). *Time of hellfire: Witch-hunt in Slovakia*. RAK.
- Avola, D., Cinque, L., Foresti, G. L., & Pannone, D. (2019). Automatic Deception Detection in RGB videos using Facial Action Units. *Proceedings of the 13th International Conference on Distributed Smart Cameras*, 1–6.
- Baltrušaitis, T., Mahmoud, M., & Robinson, P. (2015). Cross-dataset learning and person-specific normalisation for automatic Action Unit detection. *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, *06*, 1–6.
- Baltrusaitis, T., Zadeh, A., Lim, Y. C., & Morency, L.-P. (2018). OpenFace 2.0: Facial Behavior Analysis Toolkit. *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, 59–66.

- Bartlett, M. S., Littlewort, G. C., Frank, M. G., & Lee, K. (2014). Automatic Decoding of Facial Movements Reveals Deceptive Pain Expressions. *Current Biology*, *24*(7), 738–743.
- Bartol, C. R., & Bartol, A. M. (2003). *Psychology and Law: Theory, Research, and Application* (3rd edition). Wadsworth Publishing.
- Baumgartner, T., Fischbacher, U., Feierabend, A., Lutz, K., & Fehr, E. (2009). The Neural Circuitry of a Broken Promise. *Neuron*, *64*(5), 756–770.
- Bhamare, A. R., Katharguppe, S., & Silviya Nancy, J. (2020). Deep Neural Networks for Lie Detection with Attention on Bio-signals. *2020 7th International Conference on Soft Computing & Machine Intelligence (ISCMI)*, 143–147.
- Bond, C. F., & DePaulo, B. M. (2006). Accuracy of Deception Judgments. *Personality and Social Psychology Review*, *10*(3), 214–234.
- Brewer, N., & Williams, K. D. (A c. Di). (2007). *Psychology and Law: An Empirical Perspective* (1st edition). The Guilford Press.
- Bryant, E. M. (s.d.). *Real Lies, White Lies and Gray Lies: Towards a Typology of Deception*. *7*, 27.
- Buller, D. B., & Burgoon, J. K. (1996). Interpersonal Deception Theory. *Communication Theory*, *6*(3), 203–242.
- Burgoon, J. K. (2019). Separating the Wheat From the Chaff: Guidance From New Technologies for Detecting Deception in the Courtroom. *Frontiers in Psychiatry*, *9*.
- Cantarero, K., Van Tilburg, W. A. P., & Szarota, P. (2018). Differentiating everyday lies: A typology of lies based on beneficiary and motivation. *Personality and Individual Differences*, *134*, 252–260.

- Cheney, D. L., & Seyfarth, R. M. (1990). *How monkeys see the world: Inside the mind of another species*. University of Chicago Press.
- Clemente, C. D. (1997). *Anatomy: A regional atlas of the human body* (4th ed.). Williams & Wilkins.
- Cohn, J. F., Ambadar, Z., & Ekman, P. (2007). Observer-based measurement of facial expression with the Facial Action Coding System. In *Handbook of emotion elicitation and assessment* (pag. 203–221). Oxford University Press.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297.
- Curci, A., Lanciano, T., Battista, F., Guaragno, S., & Ribatti, R. M. (2019). Accuracy, Confidence, and Experiential Criteria for Lie Detection Through a Videotaped Interview. *Frontiers in Psychiatry*, *9*.
- Darwin, C., & Prodger, P. (1998). *The Expression of the Emotions in Man and Animals*. Oxford University Press.
- De Boulogne, G.-B. D. (1862). *Mécanisme de la physionomie humaine: Atlas*. Renouard.
- Denault, V., & Dunbar, N. (2019). *Credibility assessment and deception detection in courtrooms: Hazards and challenges for scholars and legal practitioners*.
- DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M., & Epstein, J. A. (1996). Lying in everyday life. *Journal of Personality and Social Psychology*, *70*(5), 979–995.
- DePaulo, B. M., & Kirkendol, S. E. (1989). The Motivational Impairment Effect in the Communication of Deception. In J. C. Yuille (A c. Di), *Credibility Assessment* (pagg. 51–70). Springer Netherlands.

- DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, *129*(1), 74–118.
- Di Battista, P. (1994). Effects of planning on performance of trust-violating versus tactful, white lies: How are familiar speech acts cognitively represented? *Communication Studies*, *45*(2), 174–186.
- Ekman, P. (1992). *Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage* (2° edizione). W W Norton & Co Inc.
- Ekman, P. (1994). Strong evidence for universals in facial expressions: A reply to Russell's mistaken critique. *Psychological Bulletin*, *115*(2), 268–287.
- Ekman, P. (1996). Why don't we catch liars? *Social research*, *63*, 801–817.
- Ekman, P. (2003). Darwin, Deception, and Facial Expression. *Annals of the New York Academy of Sciences*, *1000*(1), 205–221.
- Ekman, P., & Friesen, W. V. (1969). Nonverbal Leakage and Clues to Deception. *Psychiatry*, *32*(1), 88–106.
- Ekman, P., & Friesen, W. V. (1978). *Facial action coding system: Manual*. Consulting Psychologists Press.
- Ekman, P., Friesen, W. V., & Hager, J. (2002). . *Facial Action Coding System [E-book]*. Research Nexus.
- Ekman, P., Roper, G., & Hager, J. C. (1980). Deliberate Facial Movement. *Child Development*, *51*(3), 886–891.
- Ekman, P., & Rosenberg, E. (2005). *What the face reveals. 2nd*. Oxford University Press.
- Ekman, P., Rosenberg, E., & Hager, J. (1998). Facial action coding system interpretive database (FACSAID). *Unpublished manuscript. University of California at San Francisco, Human Interaction Laboratory*.

- Fernandez-Duque, D., Baird, J. A., & Posner, M. I. (2000). Executive Attention and Metacognitive Regulation. *Consciousness and Cognition*, *9*(2), 288–307.
- Ford, E. B. (2006). Lie detection: Historical, neuropsychiatric and legal dimensions. *International Journal of Law and Psychiatry*, *29*(3), 159–177.
- Frank, M. G., & Ekman, P. (1997). The ability to detect deceit generalizes across different types of high-stake lies. *Journal of Personality and Social Psychology*, *72*(6), 1429–1439.
- Frank, M. G., & Svetieva, E. (2012). Lies worth catching involve both emotion and cognition. *Journal of Applied Research in Memory and Cognition*, *1*(2), 131–133.
- Frank, M. G., & Svetieva, E. (2015). Microexpressions and Deception. In M. K. Mandal & A. Awasthi (A c. Di), *Understanding Facial Expressions in Communication* (pagg. 227–242). Springer India.
- Frank, M. G., Svetieva, E., Hurley, C. M., Kang, S., Sweet, D., Pazian, M., & Ekman, P. (2014). Detecting lies in a counter terrorism situation: Body Language and Incongruity. *Manuscript in preparation*.
- Freud, S. (1959). *Collected papers* (Vol. 3). Basic Books.
- Goffman, E. (1974). *Frame analysis: An essay on the organization of experience*. Harper & Row.
- Gombos, V. A. (2006). The Cognition of Deception: The Role of Executive Processes in Producing Lies. *Genetic, Social, and General Psychology Monographs*, *132*(3), 197–214.
- Greene, J. D., & Paxton, J. M. (2009). Patterns of neural activity associated with honest and dishonest moral decisions. *Proceedings of the National Academy of Sciences*, *106*(30), 12506–12511.

- Gunnery, S. D., & Hall, J. A. (2014). The Duchenne smile and persuasion. *Journal of Nonverbal Behavior*, *38*(2), 181–194.
- Hasan, K., Rahman, W., Gerstner, L., Sen, T., Lee, S., Haut, K. G., & Hoque, M. (2019). Facial Expression Based Imagination Index and a Transfer Learning Approach to Detect Deception. *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*, 634–640.
- Hopper, R., & Bell, R. A. (1984). Broadening the deception construct. *Quarterly Journal of Speech*, *70*(3), 288–302.
- Hurley, C. M. (2012). Do you see what I see? Learning to detect micro expressions of emotion. *Motivation and Emotion*, *36*(3), 371–381.
- Hurley, C. M., & Frank, M. G. (2011). Executing facial control during deception situations. *Journal of Nonverbal Behavior*, *35*(2), 119–131.
- Ishihara, L., & Brayne, C. (2006). What is the evidence for a premorbid parkinsonian personality: A systematic review. *Movement Disorders*, *21*(8), 1066–1072.
- Izard, C. E. (1977). *Human Emotions*. plenum.
- Izard, C. E. (1994). Innate and universal facial expressions: Evidence from developmental and cross-cultural research. *Psychological Bulletin*, *115*(2), 288–299.
- Jiabo, K. (2021). *DeepLie: Detect Lies with Facial Expression (Computer Vision).*
- Jordan, S., Brimbal, L., Wallace, D. B., Kassin, S. M., Hartwig, M., & Street, C. N. H. (2019). A test of the micro-expressions training tool: Does it improve lie detection? *Journal of Investigative Psychology and Offender Profiling*, *16*(3), 222–235.

- Jupe, L. M., & Keatley, D. A. (2020). Airport artificial intelligence can detect deception: Or am i lying? *Security Journal*, *33*(4), 622–635.
- Kahn, E. A. (1966). On facial expression. *Neurosurgery*, *12*(CN_suppl_1), 9–22.
- Kassin, S. M. (2005). On the Psychology of Confessions: Does Innocence Put Innocents at Risk? *American Psychologist*, *60*(3), 215–228.
- Keltner, D. (1995). Signs of appeasement: Evidence for the distinct displays of embarrassment, amusement, and shame. *Journal of Personality and Social Psychology*, *68*(3), 441–454.
- Keltner, D., & Buswell, B. N. (1997). Embarrassment: Its distinct form and appeasement functions. *Psychological Bulletin*, *122*(3), 250–270.
- Kennedy, P. (2014). *Artificial intelligence lie detector developed by Imperial alumnus | Imperial News | Imperial College London*. Imperial College London.
- Krishnamurthy, G., Majumder, N., Poria, S., & Cambria, E. (2018). *A Deep Learning Approach for Multimodal Deception Detection* (arXiv:1803.00344). arXiv.
- Krumhuber, E. G., & Manstead, A. S. R. (2009). Can Duchenne smiles be feigned? New evidence on felt and false smiles. *Emotion (Washington, D.C.)*, *9*(6), 807–820.
- La Tona, G., Terranova, M. C., Vernuccio, F., Lo Re, G., Salerno, S., Zerbo, S., & Argo, A. (2020). Lie Detection: FMRI. In G. Lo Re, A. Argo, M. Midiri, & C. Cattaneo (A c. Di), *Radiology in Forensic Medicine: From Identification to Post-mortem Imaging* (pagg. 197–202). Springer International Publishing.
- Lewis, J. A., & Cuppari, M. (2009). The Polygraph: The Truth Lies within. *Journal of Psychiatry and Law*, *37*(1), 85–92.

- Lewis, M. (1993). The development of deception. In *Lying and deception in everyday life* (pagg. 90–105). The Guilford Press.
- MacDonald, A. W., Cohen, J. D., Stenger, V. A., & Carter, C. S. (2000). Dissociating the Role of the Dorsolateral Prefrontal and Anterior Cingulate Cortex in Cognitive Control. *Science*, *288*(5472), 1835–1838.
- Mahesh, B. (2018). *Machine Learning Algorithms—A Review*. *9*(1), 7.
- Majumdar, D. M. A., & Mostakim, M. M. (2018). *Using Machine Learning for Lie Detection: Classification of Human Visual Morphology*. BRAC University.
- Martin, C. W. (A c. Di). (2009). *The philosophy of deception*. Oxford University Press.
- Matsumoto, D. (2009). *The Cambridge dictionary of psychology*. Cambridge University Press.
- Matsumoto, D., Hwang, H. C., Skinner, L. G., & Frank, M. G. (2014). Positive Effects in Detecting Lies from Training to Recognize Behavioral Anomalies. *Journal of Police and Criminal Psychology*, *29*(1), 28–35.
- McCornack, S. A. (1997). The generation of deceptive messages: Laying the groundwork for a viable theory of interpersonal deception. In *Message production: Advances in communication theory* (pagg. 91–126). Lawrence Erlbaum Associates Publishers.
- Menza, M. (2000). The personality associated with Parkinson's disease. *Current Psychiatry Reports*, *2*(5), 421–426.
- Miehlke, A. (1973). *Surgery of the facial nerve*. PN: Saunders.
- Mohamed, F. B., Faro, S. H., Gordon, N. J., Platek, S. M., Ahmad, H., & Williams, J. M. (2006). Brain Mapping of Deception and Truth Telling about an

Ecologically Valid Situation: Functional MR Imaging and Polygraph Investigation—Initial Experience. *Radiology*, *238*(2), 679–688.

- Myers, R. E. (1976). Comparative Neurology of Vocalization and Speech: Proof of a Dichotomy. *Annals of the New York Academy of Sciences*, *280*(1), 745–757.
- Oravec, J. A. (2022). The emergence of “truth machines”?: Artificial intelligence approaches to lie detection. *Ethics and Information Technology*, *24*, 1–10.
- Oster, H. (2001). *Baby FACS: Facial Action Coding System for infants and young children*.
- Oster, H., Camras, L., Campos, J., Campos, R., Ujiee, T., Zhao-Lan, M., & Lei, W. (1996). *The patterning of facial expressions in Chinese, Japanese, and American infants in fear-and anger-eliciting situations*.
- Owayjan, M., Kashour, A., Al Haddad, N., Fadel, M., & Al Souki, G. (2012). The design and development of a Lie Detection System using facial micro-expressions. *2012 2nd International Conference on Advances in Computational Tools for Engineering Applications (ACTEA)*, 33–38.
- Pérez-Rosas, V., Abouelenien, M., Mihalcea, R., & Burzo, M. (2015). Deception Detection using Real-life Trial Data. *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, 59–66.
- Peskin, J. (1992). Ruse and Representations: On Children’s Ability to Conceal Information. *Developmental Psychology - DEVELOP PSYCHOL*, *28*, 84–89.
- Pickel, K. (2004). When a lie becomes the truth: The effects of self-generated misinformation on eyewitness memory. *Memory*, *12*(1), 14–26.
- Porter, S., & ten Brinke, L. (2008). Reading Between the Lies: Identifying Concealed and Falsified Emotions in Universal Facial Expressions. *Psychological Science*, *19*(5), 508–514.

- Porter, S., ten Brinke, L., & Wallace, B. (2012). Secrets and Lies: Involuntary Leakage in Deceptive Facial Expressions as a Function of Emotional Intensity. *Journal of Nonverbal Behavior*, *36*(1), 23–37.
- Rafter, N. (2005). The murderous Dutch fiddler: Criminology, history and the problem of phrenology. *Theoretical Criminology*, *9*(1), 65–96.
- Rill-Garcia, R., Jair Escalante, H., Villasenor-Pineda, L., & Reyes-Meza, V. (2019). *High-Level Features for Multimodal Deception Detection in Videos*. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops.
- Rowe, J. B., Toni, I., Josephs, O., Frackowiak, R. S. J., & Passingham, R. E. (2000). The Prefrontal Cortex: Response Selection or Maintenance Within Working Memory? *Science*, *288*(5471), 1656–1660.
- Saxe, L. (1991). Lying: Thoughts of an applied social psychologist. *American Psychologist*, *46*(4), 409–415.
- Schönfeld, V. (2007). *Textbook of scientific graphology for beginners*. Elfa Press.
- Seiter, J. S., Bruschke, J., & Bai, C. (2002). The acceptability of deception as a function of perceivers' culture, deceiver's intention, and deceiver-deceived relationship. *Western Journal of Communication*, *66*(2), 158–180.
- Sporer, S. L., & Schwandt, B. (2006). Paraverbal Indicators of Deception: A Meta-analytic Synthesis. *Applied Cognitive Psychology*, *20*(4), 421–446.
- Sporer, S. L., & Schwandt, B. (2007). Moderators of nonverbal indicators of deception: A meta-analytic synthesis. *Psychology, Public Policy, and Law*, *13*(1), 1–34.
- Sternglanz, R. W., Morris, W. L., Morrow, M., & Braverman, J. (2019a). A Review of Meta-Analyses About Deception Detection. In T. Docan-Morgan (A c.

Di), *The Palgrave Handbook of Deceptive Communication* (pagg. 303–326). Springer International Publishing.

- Sternglanz, R. W., Morris, W. L., Morrow, M., & Braverman, J. (2019b). A Review of Meta-Analyses About Deception Detection. In T. Docan-Morgan (A c. Di), *The Palgrave Handbook of Deceptive Communication* (pagg. 303–326). Springer International Publishing.
- Su, L., & Levine, M. D. (2014). High-Stakes Deception Detection Based on Facial Expressions. *2014 22nd International Conference on Pattern Recognition*, 2519–2524.
- Sullivan, E. (2001). *The concise book of lying*. Farrar, Straus and Giroux Press.
- Takabatake, S., Shimada, K., & Saitoh, T. (2018). Construction of a Liar Corpus and Detection of Lying Situations. *2018 Joint 10th International Conference on Soft Computing and Intelligent Systems (SCIS) and 19th International Symposium on Advanced Intelligent Systems (ISIS)*, 971–976.
- Takalkar, M., Xu, M., Wu, Q., & Chaczko, Z. (2018). A survey: Facial micro-expression recognition. *Multimedia Tools and Applications*, *77*(15), 19301–19325.
- Tambe, P., Cappelli, P., & Yakubovich, V. (2019). Artificial Intelligence in Human Resources Management: Challenges and a Path Forward. *California Management Review*, *61*(4), 15–42.
- *The expression of the emotions in man and animals*. (s.d.). 399.
- Tooby, J., & Cosmides, L. (1989). The logic of threat. *Human Behavior and Evolution Society, Evanston, IL.*
- Trovillo, P. V. (1938). History of Lie Detection Police Science. *Journal of the American Institute of Criminal Law and Criminology*, *29*(6), 848–881.

- Tschiassny, K. (1953). LX Eight Syndromes of Facial Paralysis and Their Significance in Locating the Lesion. *Annals of Otology, Rhinology & Laryngology*, *62*(3), 677–691.
- Turner, R. E., Edgley, C., & Olmstead, G. (1975). INFORMATION CONTROL IN CONVERSATIONS: HONESTY IS NOT ALWAYS THE BEST POLICY. *The Kansas Journal of Sociology*, *11*(1), 69–89.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is All You Need. *Advances in Neural Information Processing Systems*, *30*.
- Vendemia, J. M. C., & Nye, J. M. (2018). The Neuroscience of Deception. In A. R. Beech, A. J. Carter, R. E. Mann, & P. Rotshtein (A c. Di), *The Wiley Blackwell Handbook of Forensic Neuroscience* (1ª ed., pagg. 171–188). Wiley.
- Venkatesh, S., Ramachandra, R., & Bours, P. (2019, March). Robust algorithm for multimodal deception detection. In *2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)* (pp. 534-537).
- Vicianova, M. (2015). Historical Techniques of Lie Detection. *Europe's Journal of Psychology*, *11*(3), 522–534.
- Vrij, A. (2000). *Detecting lies and deceit: The psychology of lying and implications for professional practice*. Wiley.
- Vrij, A. (2009). *Detecting lies and deceit: Pitfalls and opportunities* (2. ed., repr). Wiley.
- Vrij, A., Mann, S., & Fisher, R. P. (2006). Information-gathering vs accusatory interview style: Individual differences in respondents' experiences. *Personality and Individual Differences*, *41*(4), 589–599.

- Walczyk, J. J., Roper, K. S., Seemann, E., & Humphrey, A. M. (2003). Cognitive mechanisms underlying lying to questions: Response time as a cue to deception. *Applied Cognitive Psychology*, *17*(7), 755–774.
- Wellman, H. M. (1992). *The child's theory of mind* (pagg. xiii, 358). The MIT Press.
- Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*(1), 103–128.
- Yan, W.-J., Wu, Q., Liang, J., Chen, Y.-H., & Fu, X. (2013). How Fast are the Leaked Facial Expressions: The Duration of Micro-Expressions. *Journal of Nonverbal Behavior*, *37*(4), 217–230.
- Yan, W. J., Wang, S. J., Liu, Y. J., Wu, Q., & Fu, X. (2014). For micro-expression recognition: Database and suggestions. *Neurocomputing*, *136*, 82-87.
- Zeki, S., Goodenough, O. R., Spence, S. A., Hunter, M. D., Farrow, T. F. D., Green, R. D., Leung, D. H., Hughes, C. J., & Ganesan, V. (2004). A cognitive neurobiological account of deception: Evidence from functional neuroimaging. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *359*(1451), 1755–1762.
- Zhou, Z. H. (2016). Learnware: on the future of machine learning. *Frontiers Comput. Sci.*, *10*(4), 589-590.
- Zloteanu, M. (2020). *Reconsidering Facial Expressions and Deception Detection* (pagg. 238–284).
- Zuckerman, M., DePaulo, B. M., & Rosenthal, R. (1981). Verbal and Nonverbal Communication of Deception. In *Advances in Experimental Social Psychology* (Vol. 14, pagg. 1–59). Elsevier.