THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l'Université Toulouse 2 - Jean Jaurès

Présentée et soutenue par

Camille SROUR

Le 7 décembre 2021

Théorie Générale du Mensonge : Une théorie disruptive de la production, prévention et détection du mensonge

---

Ecole doctorale : **CLESCO - Comportement, Langage, Education, Socialisation, Cognition**

Spécialité : **Psychologie**

Unité de recherche :
CLLE - Unité Cognition, Langues, Langage, Ergonomie

Thèse dirigée par
Jacques PY

Jury

Mme Astrid HIRSCHELMANN, Rapporteure
M. Colin TREDOUX, Rapporteur
M. Jaume MASIP, Examinateur
M. Samuel DEMARCHI, Examinateur
M. Jacques PY, Directeur de thèse

Table of Contents

List of Figures

List of Tables

Acknowledgements

To my supervisor, Professor Jacques Py, for his trust and the unique autonomy he offered me, without which I probably never would have attempted to build such a thing as a General Theory of Deception.

To the jury of my defence: Dr Samuel Demarchi, Professor Astrid Hirschelmann, Professor Jaume Masip, and Professor Colin Tredoux for their interest in my work, the time taken to read this thesis, our exchanges along the way, and their presence at my defense.

To Professor Aldert Vrij, for accepting a decade ago to welcome a young engineer in his lab with no academic background in psychology at the time. My passion for the field really started in Portsmouth.

To Sharon Leal, for her training in deception detection, and our exchanges and even debates over the years on the most efficient interviewing techniques and cues to deceit.

Finally, thanks to you, whoever you are, for reading this thesis. I hope you will enjoy it as much as I *sincerely* enjoyed conducting this research, writing about it, and exploring the means to *making people more honest*!

Abstract

The General Theory of Deception (GTD) aims to unify and complete the various sparse theoretical units that have been proposed in the deception literature to date, in a comprehensive framework fully describing from end to end how and when deceptive messages are produced, and how this can inform more effective prevention and detection.

As part of the elaboration of the theory, the different ways people elaborate deceptive messages were first tracked by the author daily, over 3 years, resulting in the identification, description, and naming of 98 “Elementary Deception Modes” (86 verbal, 12 nonverbal) that can all be combined during one deceptive episode, thus leading to a total estimate of 10^{29} different ways to lie.

Central to the GTD is the "Five Forces Model", explaining precisely at which times deceptive messages occur and what factors compete to determine the types of messages that are most likely to be produced (truthful, refusal to answer, or deceptive – and with which deception modes). Finally, the process by which deceptive messages come to mind and are compared, both against each other and against the option of disclosing the truth, given memory's capacity and time limits, has been described in the form of a dynamic, continuous, and testable algorithm called the “Deception Decision Algorithm” (DDA).

The practical insights derived from this new disruptive theory of lie production are discussed and a theory-based lie prevention and detection enhancement method is introduced. Finally, three series of experiments were carried out to test certain predictions of the theory, in particular the impact of the manipulation of factors within the Five Forces Model on the prevention and enhancement of deception detection, experimentally confirming the main predictions for practical applications of the GTD theory.

Résumé

La Théorie Générale du Mensonge (General Theory of Deception, GTD) vise à unifier et à compléter les unités théoriques éparses qui ont été proposées dans la littérature sur le mensonge à ce jour. A travers une théorie complète décrivant de bout en bout quand et comment les messages trompeurs sont produits, celle-ci fournit des clés pour une prévention et une détection du mensonge plus efficaces.

Dans le cadre de l'élaboration de la théorie, les différentes manières dont tout un chacun élabore des messages trompeurs ont d'abord été suivies quotidiennement, sur une période de 3 ans, ce qui a permis d'identifier, de décrire et de nommer 98 « modes de mensonge élémentaires » (86 verbaux, 12 non verbaux), pouvant tous être combinés au cours d'un épisode mensonger, et conduisant ainsi à une estimation totale de 10^{29} façons différentes de mentir.

Elément central de la théorie GTD, le « Modèle des Cinq Forces » (Five Forces Model) explique précisément quand et comment les messages trompeurs sont générés et quels facteurs entrent en concurrence pour déterminer les types de messages les plus susceptibles d'être produits (véridique, refus de répondre ou mensonger - et avec quels modes de mensonge). Enfin, le processus par lequel les messages trompeurs viennent à l'esprit et sont comparés, à la fois entre eux et par rapport à l'option de révéler la vérité, compte tenu des limites temporelle et de capacité de la mémoire, a été décrit sous la forme d'un algorithme dynamique, continu et testable appelé « Algorithme de Décision de Mensonge » (Deception Decision Algorithm, DDA).

Les applications pratiques et les prédictions découlant de cette nouvelle théorie disruptive de la production de mensonges sont discutées, et une méthode d'amélioration de la prévention et de la détection des mensonges fondée sur cette théorie est présentée. Enfin, trois séries d’expérimentations ont été conduites pour tester certaines prédictions de la théorie, notamment

l'impact de la manipulation de facteurs du Modèle des Cinq Forces sur la prévention et l'amélioration de la détection du mensonge, confirmant expérimentalement les grandes prédictions à visée pratique de la théorie GTD.

General Introduction

Knowing that I had been studying deception detection for several years, a former classmate, now working as a professional buyer at a major international dairy company, contacted me for some advice on his daily negotiations. His job consisted of trying to buy plastic (e.g., for their yoghurt cups) at the best possible price to cover their quarterly needs. Because plastic is derived from oil, its market price fluctuates daily. When the prices were high, he would buy the minimum volumes of plastic he needed, when they were low, he would buy more and stock up. With his few suppliers on the market, most negotiations were done by email, and our professional buyer always doubted he was getting the best price from the suppliers, or if they all somehow agreed on favorable prices for them. To detect potential cues of deception, I asked to see some of their typical email exchanges; however, they contained such short statements there were no cues to deceit to detect. For example, a typical email laconically said, "Our offer is €915 per ton of plastic, thank you".

Most credibility assessment tools described in the literature, like Criteria Based Content Analysis (CBCA; Köhnken & Steller, 1988; Raskin & Esplin, 1991; Raskin & Steller, 1989; Raskin & Yuille, 1989; Steller, 1989; Steller & Boychuk, 1992; Steller & Köhnken, 1989; Yuille, 1988) or Reality Monitoring (RM; Johnson & Raye, 1981; Masip et al., 2005; Sporer, 2004), rely on long and open narratives from interviewees to be able to discriminate liars from truth tellers above the level of chance. Yet, these are not applicable to the simple, but most common, daily situations where deception could take place. More recently, researchers have advocated moving on from an attempt to passively observe behavioral cues to deception in an interviewee's statement (Buller et al., 1994, 1996; Ekman, 1985; Ekman & Friesen, 1969; Vrij et al., 2006), to actively adopting interviewing strategies designed to magnify the behavioral

differences between liars and truth tellers in order to better discriminate them. For example, this line of research showed that asking cognitively demanding questions to interviewees (e.g., recounting activities in reverse chronological order, asking unexpected questions like drawing one's activities, answering as quickly as possible, etc.) was judged more difficult by both liars and truth tellers, but much more so by liars (Vrij, Mann et al., 2008, 2009, 2012; Walczyk et al., 2005, 2009, 2012, 2013). This generally resulted in more salient cues to deception (e.g., quantity of details), improving the classification of liars vs. truth tellers, as compared to a passive interviewing classification attempt.

As outlined above, however, asking for open narratives from interviewees, especially when combined with strategic interviewing instructions, can be inapplicable, inappropriate, or simply judged too time-consuming in non-forensic (most common) interviewing settings. For example, these strategies aim to determine whether the description of past activities has been fabricated, and do not apply to our plastic buyer, as our professional needs to evaluate the veracity of a factual statement (in this case, the price). Similarly, in a job interview context, if a recruiter could ask a candidate to describe their typical weekly tasks at each of their past professional experiences in as much detail as possible (i.e., asking for an open narrative) to determine if they genuinely worked at the companies in question, it still would not help determine the crucial question of the candidate's current salary. Finally, even in forensic contexts where suspects deny any wrongdoing rather than having to describe activities (e.g., "I do not, I have never had, an illegal bank account in Switzerland"), traditional credibility assessment tools become of little use to professionals.

To address such common yet understudied real-life deception, a new approach to credibility evaluation is proposed here. After passive deception detection and active strategic

interviewing for better deception detection, we also propose techniques that try to prevent people from lying in the first place. For deception about facts in particular, or false denials, where no trace of deception can be detectable, an alternative approach is to try to deter people from perceiving lying as their best option. But what exactly makes people think lying is their best option, when solicited for the truth? What psychological factors, ingredients, and "recipe" do we all mentally perform when asked for the truth in order to choose whether to lie, and if so, using which deceptive strategy? If such factors exist and can be reliably identified, could they be manipulated to decrease the likelihood of people acting or answering deceitfully? And even if they do, could manipulating these factors at least make the cues to deceit more salient?

As the core of the present work, a General Theory of Deception (GTD) is first proposed. It aims to unify and complete the various sparse theoretical units that have been proposed in the deception literature to date, in a comprehensive psychological model of deception, fully describing from end to end how and when deceptive messages are produced, and how this can inform more effective prevention and detection. The elaboration of the theory started with a 3-year daily observation, recording, and description of the different ways people elaborate deceptive (verbal and nonverbal) messages in real life, which we called the "deception modes". Central to the GTD is the "Five Forces Model", explaining precisely at which times deceptive messages occur and describing the five factors that compete to determine the types of messages that are most likely to be produced (truthful, refusal to answer, or deceptive), and with which deception modes. Finally, the process by which deceptive messages come to mind and are compared, both against each other and against the option of disclosing the truth, given one's memory capacity and time limits, has been described in the form of a dynamic, continuous, and testable algorithm called the "Deception Decision Algorithm" (DDA).

A series of experiments were then conducted to test the theory's predictions. As the initial motivation to elaborate the GTD, an online experiment was set up simulating a buyer/supplier email negotiation about the price of plastic for the packaging of dairy goods. A total of 458 participants were recruited to act as suppliers, and divided into eight experimental groups that only differed in one sentence of the first email they received from the buyer asking for a quote. All participants, as suppliers, were ultimately asked for their own purchasing price by the buyer – a highly sensitive piece of information. Two out of the five GTD factors (the Dissonance and Risk factors) were manipulated between the experimental groups to analyse their effects on the prevention of deceitful information disclosure (their purchasing price of plastic as suppliers) by the participants.

Next, the application of GTD to real-life high-stakes situations was performed in the context of recruitment. By far the most complex to put in place (almost 4 years in total from initial discussions to data analysis), I collaborated for this field experiment with a recruiting firm (collecting candidates' applications) and a background check firm (performing background checks to verify the veracity of resume claims) to test a mechanism that addresses resume fraud in recruitment applications. Resume deception could contain stretched dates, boosted job titles, or fabricated academic degrees. We designed a customized application form to both *deter* and *enhance* detection of deception in applications by manipulating three GTD factors (Punishment, Dissonance, and Risk). Twenty-seven real-life candidates all applying for sales positions were asked to complete the customized application form at the pre-interview stage, after submitting their initial free-form resume. A total of 269 elements common to both application forms were verified, and the extent to which the customized application form diminished the proportion of

deceptive statements from candidates was analysed. The ability of the customized application form to make any remaining deception in applications more detectable was also tested.

While the two previous experimental studies tested the GTD for deception prevention (experiment 1, online negotiation) and prevention + detection (experiment 2, recruitment field study), our last experiment focused on the manipulation of a GTD factor, the Execution cost, for the sole purpose of deception detection enhancement. In line with the literature on increasing cognitive load as a means to magnify the cues to deceit, we conducted a forensic-like scenario where participants stole, or not, a 10-euro store voucher, and were subsequently all interviewed about their whereabouts during the previous 20 minutes. Participants were randomly assigned to one of four experimental groups, examining the optimal instructions of a within-subject two-recall strategy to detect deception during an interview. In other words, this experiment compared different interviewing instructions to determine which manipulations of cognitive load are most effective for detecting deception.

Collectively, this work addresses whether prevention and detection of deception can be improved based on a theory-based framework.

Chapter 1: The General Theory of Deception: A Disruptive Theory of Lie Production, Prevention, and Detection.

The General Theory of Deception (GTD) aims to unify and complete the various sparse theoretical units that have been proposed in the deception literature to date, in a comprehensive framework fully describing from end to end the process by which deceptive messages are produced, and how this can inform more effective prevention and detection. As part of the elaboration of the theory, the different ways people elaborate deceptive messages were first tracked by the authors daily, over 3 years, resulting in the identification, description, and naming of 98 “Elementary Deception Modes” (86 verbal, 12 nonverbal) that can all be combined during one deceptive episode, thus leading to a total estimate of 10^{29} different ways to lie. Central to the GTD is the "Five Forces Model", explaining precisely at which times deceptive messages occur and what factors compete to determine the types of messages that are most likely to be produced (truthful, refusal to answer, or deceptive – and with which deception modes). Finally, the process by which deceptive messages come to mind and are compared, both against each other and against the option of disclosing the truth, given memory's capacity and time limits, has been described in the form of a dynamic, continuous, and testable algorithm called the “Deception Decision Algorithm” (DDA). The practical insights derived from this new disruptive theory of lie production are discussed and a theory-based lie prevention and detection enhancement method is introduced.

A Traditional Focus on the Cues Rather than the Production of Deceit

Research on deception has traditionally focused on behavioral cues that would allow practitioners to assess the credibility of statements and distinguish liars from truth tellers. For example, the Four-Factor Theory of deception (Zuckerman et al., 1981) proposes that during deceptive discourse, liars may feel increased (a) general arousal, (b) emotional load (e.g., guilt or fear), (c) cognitive load, and (d) signs of control and impression management to appear honest. While research in this area historically focused on the emotional/arousal approach (e.g., Buller et al., 1994, 1996; Ekman, 1985; Ekman & Friesen, 1969; Vrij et al., 2006), in recent years this focus has shifted to the cognitive approach, with much of the current deception research investigating cognitive interviewing strategies to maximize behavioral differences between liars and truth tellers (Walczyk et al., 2013).

According to the cognitive approach proponents, lying is more cognitively demanding than truth telling (Sporer & Schwandt, 2006; Vrij, Fisher et al., 2008; Zuckerman et al., 1981). As such, proposed interview strategies include, for example, imposing cognitive load by asking suspects to recall their stories in reverse order (Vrij, Mann et al., 2008, 2012), instructing examinees to answer as quickly as possible (Walczyk et al., 2005, 2009, 2012), or asking unexpected questions such as asking suspects to draw their responses (Vrij et al., 2009). These strategies have indeed proven to be effective in many cases (Walczyk et al., 2013). However, the theoretical basis on which they rely remains unclear. Also, as recently pointed out in the Information Manipulation Theory 2 (McCornack et al., 2014), most research in the area assumes what the authors call a "bald-faced lie (BFL) vs. bald-faced truth (BFT) dichotomy", with most experimental paradigms to date having participants either tell a 100% true or 100% false statement. Yet, even a quick glance at the way people actually produce deceptive statements in

real life shows that the "BFL vs. BFT" paradigm is not an accurate representation of everyday interactions; and BFL fabrications are an extreme and rare occurrence. In 1993, for example, French President François Mitterrand was accused of having ordered illegal wiretapping during his first mandate (which was later confirmed to be the case). Documentary footage shows that when questioned by a journalist about those accusations, he declared:

> The Elysée [French equivalent of the White house] is not wiretapping anything. There is no wiretapping system here. And I, personally, have never read any of them…There is no wiretapping service at the Elysée. There cannot be. I do not know, by the way, how to perform wiretapping. (Le Paige, 2011).

This excerpt will be dissected later in this article, but it is already clear that the deception was more subtle than pure fabrication BFL. There is thus a need for a solid general theory capturing the process by which deceptive messages are produced. Yet, very few deception production theories have been proposed so far – and only very recently (McCornack et al., 2014; Walczyk et al., 2014). This essay thus aims to fill this theoretical gap by proposing a General Theory of Deception ("GTD"), unifying the multiple theoretical constructs that have sparsely been proposed so far, and proposing from end-to-end a model that explains and when lies are produced, what they consist of, and how to better prevent and detect them.

Existing theoretical constructs inspiring GTD

Numerous theoretical and empirical manuscripts have inspired and influenced GTD. The main works and their implications for deception production and detection are summarized and discussed below.

Information Manipulation Theory 2 (IMT2)

Information Manipulation Theory 2 (IMT2; McCornack et al., 2014) is a theory of deceptive discourse production, developed as a refinement of the original Information Manipulation Theory (IMT; McCornack, 1992). IMT was a first attempt to account for deceptive messages that did not fall under the traditional research paradigm of "BFL vs. BFT dichotomy". IMT suggested that when deceiving others, people manipulate information by *covertly* violating at least one of the four communication norms that govern rational human discourse, as outlined in the Cooperative Principle (Grice, 1989): Quantity (the amount of relevant information that is shared), Quality (the veracity of shared information), Manner (the way in which disclosed information is expressed), and Relation (the relevance of disclosed information).

However, as McCornack himself later declared, "[IMT] was not a theory. No formal, testable, falsifiable propositions were posited. Second, no account was offered regarding the production mechanisms underlying information manipulation" (McCornack et al., 2014, p. 351). IMT2 was thus later developed to address these points and build a theory based on the literature from different research fields, from linguistics to artificial intelligence, cognitive neuroscience, and speech production.

IMT2 consists of a central premise and 11 propositions grouped into three propositional sets. The central premise of IMT2 is formulated as follows:

> Deceptive and truthful discourse both are output from a speech production system involving parallel-distributed-processing guided by efficiency, memory, and means-ends reasoning; and this production process involves a rapid-fire series of cognitive cycles (involving distinct modules united by a conscious workspace), and modification of

> incrementally-constructed discourse during the turn-at-talk in response to dynamic current-state/end-state discrepancies. (McCornack et al., 2014, p. 362)

In other words, first, deceptive and truthful discourse "both are output from the same speech production system" – and not from distinct production systems, explaining why cues to deceit are subtle (DePaulo et al., 2003) and often only observable in the very specific contexts of BFLs, involving more cognitive load from liars. Second, IMT2 considers this production to be parallel-distributed-processing, as opposed to the prevailing literature which modeled deceptive speech production with an initial conscious decision to lie, followed sequentially by BFL fabrication preceding speech production. Last, this whole process is performed incrementally during speech, as an attempt to provide the best answer to a problematic prompt. This problem-solving task is dynamically guided by accessing the most easily available information in memory. A consequence of this incremental problem-solving process is that people will commonly alternate between truth and deceit during the course of a speech, as opposed to mentally creating fully deceptive messages first and then articulating them.

Activation-Decision-Construction-Action Theory (ADCAT)

According to Activation-Decision-Construction-Action Theory (ADCAT)'s authors, "IMT2 details when and why information sharing can be deceptive and explains the spontaneous lying of normal conversation but is less applicable to high-stakes lies (S. McCornack, personal communication, January 7, 2013)" (Walczyk et al., 2014, p. 24). ADCAT thus attempts to provide a cognitive account of serious high-stakes lies in response to a solicitation of truth. Such situations include "a guilty suspect interrogated by the police, an unqualified job candidate interviewed for a coveted position, or an unfaithful spouse confronted by his wife" (Walczyk et al., 2014, p. 24).

ADCAT is built around four components, hypothesized to all occur at some point in most high-stakes lies:

1. *The Activation Component* represents any aspect that causes respondents to understand that a truth is solicited by targets. These truth solicitation signals can be implicit (e.g., a police officer pulling over a car driver) or explicit (a question asked during a police interview), and are supposed to activate truths in memory, if accessible.
2. *The Decision Component* encompasses the contextual elements persuading respondents to deceive in a particular way or reminding them of a decision made previously. ADCAT postulates that the choice of whether and how to be deceptive is made based on the evaluation of the expected value of a decision ("EV"), which is calculated by $\sum\ p_i v_i$, where p_i is the probability of a particular outcome, i, multiplied by v_i, the desirability of that outcome. ADCAT posits that when truth is solicited in high-stakes situations, respondents compare the EV of their different options (e.g., tell the truth, understate the truth, totally falsify the truth, etc.) and end up choosing the option with the highest EV.
3. *The Construction Component* corresponds to the method of manipulation of information adopted to deceive the targets (e.g., falsify, exaggerate, or omit information). ADCAT anticipates that the truth or related authentic information of episodic memory should be preferred and recalled to the episodic buffer as the basis of a target's view, followed by alteration.
4. *The Action Component* is the action of liars delivering lies to targets. When the stakes are high, both liars and truth tellers generally attempt to appear sincere and relaxed,

but liars take their credibility less for granted and may need to hide signs of guilt or shame.

Because the *Decision Component* is crucial in ADCAT, and because the GTD presented here models the decision to lie or tell the truth differently, we reproduce below, for the sake of comparison, the example given in ADCAT.

> Suppose a study is conducted testing this account with actual lying. An 11-point Likert scale is used to assess the subjective utility of possible outcomes (-5 = *extremely undesirable* to respondents, 0 = *neutral*, 5 = *extremely desirable*). Consider the example of a spouse who is deciding whether to admit to a recent infidelity he had with a friend of his wife in anticipation of being confronted by her on his whereabouts on the night of the infidelity.... A one-time event he regrets, his goal is to stay happily married. In evaluating the truth option, he intuitively estimates about a 90% chance of divorce if he discloses his tryst. On the other hand, even if the marriage survives, trust and intimacy will be gone.
>
> *EV(truth telling)* = .9 (-5 : *divorce*) + .1 (-1 : *together without intimacy*) = -4.6
>
> Because of the truth's negative EV, he first considers understating it by telling his wife he slept with a stranger he met at a bar while drunk that meant nothing emotionally. He infers that this would lessen slightly the probability of divorce but still yield an unacceptable EV.
>
> *EV(understating truth)* = .7 (-5: *divorce*) + .3 (-1 : *together without trust*) = -3.8

> Finally, he infers that, because of how negatively his wife would likely receive any news of infidelity and the terrible consequences on their relationship, only a complete fabrication can achieve his goal. In evaluating this option, he intuitively estimates an 80% chance that the tryst and lie will go undetected in the long run but knows he will feel guilt at not confessing, thus a subjective utility rating of 4 rather than 5.
>
> *EV(denial false)* = .8 (4: *stay married feeling guilty*) + .2 (-5 : *divorce*) = 2.2
>
> Because it has the largest EV, the denial and false alibi are chosen (Walczyk et al., 2014, p. 27).

Our own theory and GTD model will show how our approach differs from the purely "expected outcome" evaluation proposed in ADCAT, and also how GTD's decision model can be actionable to prevent and lower the likelihood of people engaging in deception in the first place.

Truth-Default Theory

Truth-Default Theory (TDT; Levine, 2014) is a theory of deception and deception detection. Whereas IMT2 is primarily a theory of deceptive discourse production, TDT focuses more on credibility assessment and deception detection accuracy. According to the author, "[TDT's propositions] are all data based, and the explanations were initially articulated so as to offer a coherent account of the existing scientific data. The theory was not made public until original research supported and replicated every major claim" (Levine, 2014, p. 2).

TDT's central concept is that people presume, without conscious reflection, that others' communication is honest. In TDT, a novel argument is that both this truth-default state and the well-documented truth-bias, i.e., the tendency to believe that another person's communication is

honest, independently of its actual honesty (Levine et al., 1999; McCornack & Parks, 1986), are functional for most of our daily interactions. This notion is demonstrated in TDT by considering the truth-lie base rate to explain the adaptive function of the truth-default state and truth-bias. TDT specifies that outside the deception lab "the prevalence of deception is much lower than the prevalence of honest communication and therefore presuming honesty leads to belief states that are typically correct" (Levine, 2014, p. 5).

Tying together various models and empirical data, TDT formulates 14 propositions, some of which are partially reproduced below:

- *Proposition 1:* most communication by most people is honest most of the time… deception is relatively infrequent, and outright lies are more infrequent still. In fact, deception must be infrequent to be effective.
- *Proposition 2:* The prevalence of deception is not normally distributed across the population. Most lies are told by a few prolific liars.
- *Proposition 3:* Most people believe most of what is said by most other people most of the time. That is, most people can be said to be truth-biased most of the time. Truth-bias results from, in part, a default cognitive state…. Truth-bias and the truth-default are adaptive both for the individual and the species. They enable efficient communication.
- *Proposition 5:* … people lie for a reason. Deception, however, is usually not the ultimate goal, but instead a means to some other ends. That is, deception is typically tactical. Specifically, most people are honest unless the truth thwarts some desired goal or goals. The motives or desired goals achieved through communication are the same for honest and deceptive communications, and deception is reserved for situations where honesty

would be ineffectual, inefficient, and/or counterproductive in goal attainment (Levine, 2014, p. 9).

The Embedded-Processes Model of Working Memory

To explain how honest and deceptive messages are elaborated, an understanding of memory's underlying mechanisms is paramount. As a matter of fact, the most frequently studied verbal credibility assessment tools, like Criteria Based Content Analysis (CBCA; Köhnken & Steller, 1988; Raskin & Esplin, 1991; Raskin & Steller, 1989; Raskin & Yuille, 1989; Steller, 1989; Steller & Boychuk, 1992; Steller & Köhnken, 1989; Yuille, 1988), emphasize the fact that accounts of truly experienced events are recalled from memory and significantly differ in their verbal content and quality from fabricated accounts ("Undeutsch hypothesis"; Steller, 1989). Much in the same vein, the Reality Monitoring tool (RM) assumes that memories of real-life experiences are encoded, among others, through sensory information (smell, taste, sound, touch, or visual details; Johnson & Raye, 1981) and as such those sensory details are more likely to be described verbally in truthful accounts (Masip et al., 2005; Sporer, 2004).

But how exactly are such perceptions encoded in one's memory? What are the underlying mechanisms occurring in memory when prompting somebody about a specific event? Although not specifically addressing the mechanisms involved in deception production, Cowan (1988, 1998, 1999) proposed an interesting model of memory which he coined "The Embedded-Processes Model of Working Memory". Working Memory refers to "the cognitive processes that retain information in an unusually accessible state, suitable for carrying out any task with a mental component" (Cowan, 1999, p. 62). Simply put, Working Memory refers to the mental processes involved when producing or understanding language, solving a problem, making a decision, or any other thought. Such tasks require that certain information be kept in mind. For

example, as Cowan (1999) illustrates perfectly, "in language comprehension, if the first word is totally forgotten by the time the second or third word is perceived, one is in bad shape. The mnemonic functions preserving information that can be used to do the necessary work collectively make up Working Memory" (pp. 62-63). In this model, Working Memory is not a separate entity from Long-Term Memory. Rather, Working Memory is "embedded" in Long-Term Memory and constitutes a moving subset of it. More specifically, when confronted with a stimulus (e.g., an object or a solicitation), relevant items from Long-Term Memory get automatically (and temporarily) activated in what Cowan calls the "Activated Memory". Activated Memory is assumed to be time-limited. The Focus of Attention, which is capacity limited, is the subset of this Activated Memory containing the information currently in conscious awareness. In Cowan's model, both voluntary processes (managed by a "Central Executive System", defining what to voluntary pay attention to) and involuntary processes activated by novel and unknown stimuli (the "Attentional Orienting System"), conjointly define and control the Focus of Attention. The Focus of Attention can also be seen as a context-specific arbitrage between concurrently activated elements in memory. As clearly described by Cowan (1999, p. 65):

> Switching lanes on a highway probably is attention demanding in that it restricts diverse types of information processing, such as those involved in conversation or ongoing thought. In contrast, navigating the vehicle automatically according to well-learned geographical cues probably would not be considered attention demanding.

Stimuli with physical features that have remained relatively unchanged over time and are of no key importance to the individual still activate some features in memory, but they do not

elicit awareness (i.e., there is habituation of orienting). On the other hand, awareness influences processing, e.g., by increasing the number of features encoded. Finally, all the activated elements can result in new associations that can now enter Long-Term Memory as new episodes. In our own GTD, we largely base our modeling of memory activation on Cowan's Embedded-Processes Model of Working Memory.

The General Theory of Deception (GTD)

How do we Lie? A Three-Year Observation of Deception Production in Low- and High-Stakes Situations

Most of the deception literature adopts one of two approaches. The first could be called a "cue-based" approach, with studies that aim to reproduce real-life contexts (e.g., suspect police interviews) and then try to find diagnostic cues to deceit (e.g., quantity of details provided). Such empirical studies focus on deception detection, and only then as a secondary objective may try to theoretically explain why certain cues emerged as deception cues. The constellations of diagnostic cues and deception detection strategies found in this line of research thus lack a common theoretical basis to understand them collectively.

The second approach is more of a "top-down" approach, and generally focuses on deception production rather than deception detection. In this approach, fundamental or theoretical mechanisms of deception production (e.g., from a brain or known psychological model perspective) are proposed, but rarely allow for the deduction of practical applications for deception prevention or detection.

Our GTD aims to propose, from end-to-end, a psychological model of deception explaining both deception production, including deducing from it a practical framework for increasing the likelihood of deception detection in practice, and also preventing the production of

deceptive messages in the first place (i.e., fostering the conditions to "make people more honest"). Throughout the present article, we define deception and lying (which we use interchangeably) as "a successful or unsuccessful deliberate attempt [intent], without forewarning, to create in another a belief which the communicator considers to be untrue" (Vrij, 2008, p. 15).

To build our GTD, we start with a bottom-up approach consisting of identifying all the ways one might try to deceive another person. We call all those ways to deceive "Deception Modes" (DM), which are a combination of elementary units or ways to deceive: the Elementary Deception Modes (EDM). This identification has never, to our knowledge, been previously exhaustively conducted. In the GTD, this listing of DM is fundamental, as we consider that only a full picture of how people might lie may allow us to (a) infer how to better detect each of these DM and (b) better prevent their production at all. The list of EDM (Supplemental material, Appendix A) is the result of a 3 year-long daily observation, annotation, and description by the first author of how people produce deceptive messages in real life, either verbally or nonverbally. The 98 EDM identified during the observation phase were then named and typical examples of each EDM have been provided for better understanding. They have also been divided between verbal and nonverbal EDM, and grouped into six categories: *Deceptive Omission* (i.e., not disclosing crucial information believed to be of interest for the solicitor), *Falsification* (stating something believed to be untrue), *Misinterpretation* (relying on the incorrect interpretation of one's words to deceive), *Distortion* (altering a true basis), *Deceptive Evasion* (not answering specifically a solicitation with the intent to deceive), and *Deceptive Diversion* (purposefully changing the topic at stake to avoid answering and mislead the solicitor). Some examples of EDM belonging to each of these categories are provided in Table 1.1.

Table 1.1

Examples of Elementary Deception Modes (EDM) among the 98 For Each Deception Category.

Category	Total # of EDM	EDM example	Description	Illustration
Deceptive Omission	13	EDM 7 - Selective Response	When several questions asked within a single sentence, choosing to answer only the question(s) for which truth can be disclosed while omitting answers to the sensitive question(s).	A 20-year-old youngster used marijuana at a Saturday night party with friends, but did not drink any alcohol. When driving back home around 2am, his car is pulled over by the police, who asks him: "have you used any drugs or drunk alcohol tonight?" The youngster quickly answers: "Officer, I haven't had a sip of alcohol!"
Falsification	29	EDM 20 – Trial Balloon	Proposing an idea involving a third party's collaboration or consent to see how they react to the idea. If the party's reactions are acceptable, the idea is executed, otherwise it is claimed to just have been an idea.	Thomas wants to go to Kevin's apartment on Friday to watch a game with all their friends, but his parents-in-law are coming to his home on Friday for dinner. Thomas tells his wife: "By the way I forgot to tell you, Kevin invited me to watch the game with everyone on Friday." Thomas' wife answers: "No way! Did you forget my parents were coming for dinner on Friday?" Thomas: "Yeah yeah I know, I was just saying that for your information."

Category	Total # of EDM	EDM example	Description	Illustration
Misinterpretation	4	EDM 46 - False irony	Telling the truth in such an exaggerated fashion (word content and/or tone of voice and/or face) that the counterpart thinks that you must be joking.	A husband asks his wife about the doubts he has been having for a long time regarding her fidelity: "Are you cheating on me? Tell me the truth. I can take it." His wife looks at him straight in the eyes, and says: "Well of course darling I am cheating on you! I even see my lover every Monday when you are travelling!" [In reality, the husband's wife does indeed see her lover every Monday.]
Distortion	20	EDM 47 - Euphemism	Understating facts or feelings or opinions about other people or things.	A new colleague has just joined Alex's department at work. After a few weeks only, Alex has formed a definitive opinion about his new colleague: he finds him arrogant, condescending and simply put: he hates him. Yet, when asked about him by other people in the department, he invariably answers: "With him? We get along okay."
Deceptive Evasion	10	EDM 68 – Implied Response	Duping a counterpart by apparently providing an answer that implies the response to their question, without	– Police officer: "Did you kill your wife?" – Suspect: "I have never hurt anybody!" [In reality, the suspect did kill his wife, and wants the police officer to deduce from his answer that not

Category	Total # of EDM	EDM example	Description	Illustration
			actually specifically answering the question.	hurting anybody includes not killing someone. Notice though that technically, the suspect could have killed his wife without actually making her suffer.]
Deceptive Diversion	10	EDM 77 – Topic Switch	Changing the topic at stake in order not to answer a sensitive question. This EDM is convenient as it later allows, if caught, justification by saying: "I actually did not comment on that topic".	A journalist is interviewing a politician: – Journalist: How do you feel about your indictment yesterday? – Politician: “Look, I'll just say this. I find it very convenient that this indictment happens right in the middle of the campaign. How do you explain that? Do you really think there is no political involvement from the government here? Come on.” [Notice how the politician changed the topic and never talked about his feelings to the indictment but diverted the discussion to accusations against the government.]

Even though this list of EDM, that resulted from our daily observations over 3 years, might not be collectively exhaustive or perfectly mutually exclusive between EDM, it represents a minimum (extended) collection of all the EDM people can use and combine to produce deceptive messages in real life.

Revisiting the François Mitterrand example, we can now analyse it with our EDM list and code the DM used in his statement:

- "The Elysée is not wiretapping anything": *Misinterpretation EDM 43 – A Priori Play on Words*, as at this stage of the statement there is ambiguity in the word "Elysée", which can either refer to the French executive power in general, or in a more restricted definition, to the residence of the French President.
- "There is no wiretapping system here": *Deceptive Evasion EDM 68 – Implied Response* + *Deceptive Diversion EDM 77 – Topic Switch*, as the journalist is supposed to infer from that answer that because "there is no wiretapping system here", there has been no illegal wiretapping ordered at all from the President. This is also a likely attempt to dodge the topic at stake, i.e., has the French President ordered illegal wiretapping from his services (as opposed to whether wiretapping is performed at the Elysée)?
- "And I, personally, have never read any of them": *Deceptive Evasion EDM 72 – Qualifying Statement*, as using the words "I, personally" could have later allowed the President to argue that he did not lie per se because, for example, other people had read the transcripts and informed him of their main content.
- "There is no wiretapping service at the Elysée. There cannot be. I do not know by the way how to perform wiretapping": *Deceptive Evasion EDM 76 – Irrelevant proof* + *Deceptive Diversion EDM 77 – Topic Switch* + *Deceptive Diversion EDM 82 – Time*

> *Buying* as, first, repeating the same argument as before appears to be an attempt to buy time in a potential attempt to find some better argument. Then saying "there cannot be [any wiretapping service at the Elysée]" is an irrelevant proof, because the lack of wiretapping service in the residence of the President is not proof that he has not ordered illegal wiretapping. Last, adding "I do not know, by the way, how to perform wiretapping" is a topic switch, as knowing whether the President knows how to conduct such wiretapping himself has nothing to do with the accusations at stake.

It thus clearly appears that this short statement contains many subtle DM, yet without resorting to any of the BFL lies mainly studied in the deception literature to date.

How Many Ways Are There to Lie? Enumeration of the Deception Modes Derived from the EDM

As a consequence of listing all observed EDM, the question of how many DM exist, i.e., how many combinations of EDM can be used within a deceptive episode, becomes relevant. Within a deceptive episode, one can indeed use only one EDM among the 98 available, but also two, three…or 98 EDM among the total 98 available.

Mathematically speaking, the issue of enumerating the number of possible DM combinations given the (minimum number of) 98 existing EDM thus becomes one of solving and finding the value of $\sum_{k=1}^{98}\binom{98}{k}$.

This sum is equal to $2^{98}-1$, which approximates to 10^{29} (or $100 * 10^9 * 10^9 * 10^9$). In other words, this comes down to saying that there are about "a hundred billion of billions of billions" of shades of deception, or ways to lie, within a deceptive episode.

When are Lies Produced?

The Five Forces Model

Now that we know how lies are a combination of one or more of the 98 EDM, one critical question remains. When do we engage in deception? And when we do, how is the selection of the best DM performed?

Contrary to previous theoretical proposals that mainly suggest a simple analysis of the "expected outcome" of lying (Masip et al., 2016; Walczyk et al., 2014), we claim that the decision to lie depends on five and only five competing factors (a process which we call the "Five Forces Model"): perceived benefits B, perceived punishment P, perceived risk R, perceived execution cost E, and perceived dissonance D. When a situation calls for an answer from an individual, either implicitly or explicitly, one can only react in one of three ways: answer truthfully, deceptively, or decide not to answer (either by remaining silent or by explicitly saying something like "I will not answer this question").

Therefore, in the GTD, we introduce "Utility functions" of lying U_L, telling the truth U_T, and not answering U_{NA}, which we could say compare the costs and benefits of those different options when truth is solicited from an individual. All those Utility functions U have the basic following form (see below for their detailed form):

$$\begin{cases} U = B - p * P - r * R - e * E - d * D \\ with: p, r, e, d \text{ the respective weights of } P, R, E, D \\ \quad and: p + r + e + d = 1 \end{cases}$$

What might first appear as a complex modeling of the factors determining the decision to lie or tell the truth deserves a few explanations.

B_L first represents the benefits perceived by an individual to resorting to lying, whereas B_T represents the perceived benefits of telling the truth. B_{NA} represents the perceived benefits of

not answering. Benefits of lying (also referred to as "deception motives" in the literature) have been examined by several authors seeking to classify them. For example, analyses of survey responses show that benefits of lying can be categorized according to the beneficiary of the lie, i.e., (a) benefits for the self, (b) benefits for others, (c) benefits of preserving a relationship (Hample, 1980; Metts, 1989). Within these categories, others have detailed the motives for lying, e.g., (a) to avoid punishment, (b) to get something, (c) to protect friends, (d) to protect one's self, (e) to win admiration, (f) to avoid social awkwardness, (g) to avoid embarrassment, (h) to maintain privacy, and (i) for power over authority (Ekman, 1991). Interestingly, in one of the few data-driven examinations of cross-cultural similarities in deception motives (Levine et al., 2016), 10 motives captured 98% of the coded responses from 387 participants coming from five countries: (a) personal transgression (e.g., hiding an infidelity), (b) economic advantage (e.g., selling defective products), (c) non-monetary personal advantage (e.g., getting a co-worker to do a disliked task), (d) social-polite (e.g., saying that a disliked gift was liked), (e) altruistic lies (e.g., a father hiding a health problem from his child), (f) self-impression management (e.g., exaggerating accomplishments to a romantic interest), (g) malicious (e.g., spreading false rumors), (h) humor-joke (pranking another), (i) pathological lies, and (j) avoidance of another person. Motives (a), (b), (c), and (g) collectively captured 70% of the coded responses pan-culturally.

On the costs side of the Utility functions, P_L, P_T, and P_{NA} respectively represent the perceived importance by an individual of the punishments /negative consequences in case of an undesirable outcome of the options of lying, telling the truth, or not answering a solicitation. For obvious ethical reasons, the effects of punishment in lie production have rarely been studied directly in the deception literature, let alone the roles of P_L and P_T, separately or independently

from the other factors, despite the widespread popular belief that a threat of punishment can act as a deterrent to lying and play a role in whether or not people decide to engage in deception. For example, Talwar et al. (2015) investigated the production of lies in young children. The experimenter played a pretext game with the children before telling them that she would need to leave the room momentarily. She then placed a target toy behind the children and told them not to turn around and look at it while she was gone. When the experimenter returned, her remarks to the children depended on the condition they were assigned to, either rewarding honesty through an External Appeal ("If you tell the truth, I will be really pleased with you") or an Internal Appeal ("It is really important to tell the truth because telling the truth is the right thing to do when someone has done something wrong") and/or mentioning punishments for having disobeyed ("if you peeked at the toy, you will be in trouble"). In this example, P_L and P_T were therefore not differentiated. Results showed that the percentage of children lying was significantly higher for children in the "Internal Appeal/Punishment" condition than for those in the "Internal Appeal/No Punishment" condition.

For adults, only one study (Frank & Ekman, 1997) to our knowledge has transposed the stakes of punishments into the laboratory. Participants were told that if judged lying they would have to sit in a darkened room and endure a sequence of 110-decibel startling blasts of white noise over the course of 1 hour (and then given a sample of it). In his autobiographical book, Ekman (2016) anecdotally reports about this experiment, "We also ran a condition […] with no punishment. Without the punishment our [facial behavior] measurements failed to distinguish lying from truthfulness" (p. 138).

Next, R_L models the perceived risk by an individual of raising suspicions or being disbelieved when lying at the time of the lie, whereas R_T represents the perceived risk of raising

suspicions or being disbelieved despite telling the truth. R_{NA} represents the perceived risk of raising suspicions when not answering a solicitation. Subjective evaluation of these risks of raising suspicions in other people requires Theory of Mind, i.e., the ability to imagine the thoughts, feelings, knowledge, or intentions of others, but also to evaluate how likely one's lies are to be believed by another person (DePaulo, 1992; Gombos, 2006; Talwar et al., 2007). Among the rare studies examining this factor, Thijssen et al. (2017) showed that a large majority of the children were deceptive in the low-risk condition of their experiment (N = 121, 74.2%), but most children refrained from deception when at risk of getting caught (69 out of 121, 57%).

E_L, E_T, and E_{NA} then represent the "execution costs" of lying, telling the truth, or not answering a solicitation, respectively. These execution costs (DePaulo & Kirkendol, 1989; Ekman, 2009; Sporer & Schwandt, 2006; Vrij et al., 2011; Zuckerman et al., 1981) can relate to cognitive (e.g., the fabrication of the lie itself), emotional (e.g., feeling guilty about lying), or control/inhibition (e.g., inhibiting body impulses related to lying in order to appear honest) costs. For example, Vrij et al. (2011) identified six reasons why lying could be more cognitively demanding than truth telling: (a) formulating a lie itself, (b) controlling one's behavior, which can itself be cognitively demanding, (c) monitoring interviewers' reactions to assess whether they are believing the lie, (d) reminding oneself to act and role play, (e) suppressing the truth while lying, and (f) activating a lie, which is more intentional and less automatic than the activation of the truth (note that we argue though that lying can also be less cognitively demanding than telling the truth in many cases, an argument we develop more in the Discussion section). For example, Van 't Veer et al. (2014) examined the role of an increase in the cognitive execution cost of lying as a deterrent to deception. Participants rolled a die three times and reported their outcomes while either under high (memorizing a string of eight letters) or low (string of two

letters) cognitive load – knowing that they would be paid the value of only one roll. Results showed that participants in the low cognitive load condition reported significantly higher outcomes on the paid roll than those in the high load condition, suggesting that a limited cognitive capacity will foster a tendency to be honest. Interestingly, no difference between the conditions was observed for the rolls that were not paid out, i.e., in GTD terms, where there was no perceived benefit of lying ($B_L = 0$).

Last, D_L, D_T, and D_{NA} represent what we call the perceived "dissonance" (in allusion to Festinger, 1957) of an individual when telling a lie, admitting to the truth, or not responding to a solicitation for truth. In this context, dissonance can first represent the mental conflict faced by people when some of their words or behaviors are inconsistent with their own past behaviors, actions, statements, ideas, or values. We call this first kind of dissonance "internal dissonance". In the GTD, dissonance also captures the cases where an individual's behaviors are in contradiction with the values, morals, or ethics of the society or social group in which they live; although these have not been internalized as their own personal values, they can still respect them. We call this second kind "external dissonance". Because external dissonance is not necessarily internalized by the individual, we predict that it should create a lower level of mental conflict than internal dissonance. This last GTD factor, or competing force of our Five Forces Model, has rarely been studied as such in the deception literature. However some evidence suggests for example, that asking children to promise to tell the truth (that is, a way of increasing the dissonance cost of lying in children) increases truth telling (Lyon et al., 2008; Talwar et al., 2002, 2004). Likewise, the police are often advised to build rapport during interviews to obtain greater cooperation from suspects (Vallano et al., 2015; Vanderhallen & Vervaeke, 2014; Walsh & Bull, 2012), representing another way of increasing the dissonance in suspects considering

lying (i.e., building rapport can make the suspect implicitly think "because we have a connection, it is a little more conflicting to me to start lying now and breach our trust").

All these "BPRED" factors (Benefit, Punishment, Risk, Execution, Dissonance) represent an individual's subjective perception and assessment of the present and the future, at the time of the assessment. In the Utility functions, those five factors are not necessarily of equal importance, depending on individuals, context, or time of the assessment. Therefore, they are weighted in the Utility functions by coefficients that capture those variations between individuals and contexts, and weight the BPRED factors accordingly.

While all Utility functions U have the same basic form described earlier, Utility functions of non-truthful answering options differ in the GTD from that of the truth in that they take into account the subjective evaluation of possible undesirable events that can occur later on, after the individual has answered non-truthfully, and all their possible justifications. We call the resulting evaluation of all the possible justifications of a non-truthful option its "Justification Power", J. Put differently, the Utility function of a deceptive answering option N is modeled as:

$$\begin{cases} U_L^N = u * \overline{U}_L^N + \sum_i prob(i) * J_L^{i\ of\ N} \\ with: u = 1 - \sum_i prob(i) \\ and: \overline{U}_L^N = B_L^N - p_L^N * P_L^N - r_L^N * R_L^N - e_L^N * E_L^N - d_L^N * D_L^N \\ and: J_L^{i\ of\ N} = max_j\left(J_L^{j\ if\ i\ of\ N}\right) \\ with: J_L^{j\ if\ i\ of\ N} = B_L^j - p_L^j * P_L^j - r_L^j * R_L^j - e_L^j * E_L^j - d_L^j * D_L^j \end{cases}$$

With $\overline{U}_L^N$ representing the BPRED evaluation of answering with deceptive option N at the time the individual believes they will need to answer the solicitation for truth (t_{answer}), $prob(i)$ the probability that undesirable event i occurs after answering deceitfully with N at t_{answer}, and $J_L^{j\ if\ i\ of\ N}$ the BPRED evaluation of justifying lying with N at t_{answer} with another lie (or lie sequence) j at $t > t_{answer}$ in case undesirable event i occurs. Therefore, $J_L^{i\ of\ N}$ represents the

best (sequence of) justification(s) the individual can think of giving later on in case things go wrong (with event i) after their initial lie N, after comparing all the possible justifications j they could think of. Finally, the overall Utility function U_L^N of lying with N, weighs all those options against their respective likelihood $prob(i)$.

The Deception Decision Algorithm (DDA)

Based on the evaluation of the BPRED factors of our Five Forces Model, we now introduce the decision rules we posit are used by an individual whenever they are solicited for the truth to determine if they are going to engage in deception or not, and if so, with which deception mode. We refer to this decision model as the "Deception Decision Algorithm" (DDA; see Appendix B).

The main principles of the DDA can be explained in less mathematical terms. First, our theory assumes that when a solicitation for truth is made to an individual, items relevant to the solicitation are automatically and progressively activated in the individual's memory, with the speed of activation depending, for example, on the "weights" of the items in memory, the accessibility of the item in memory, and the individual's memory capacity at the time of the solicitation. Of the relevant items activated, some are specific to the solicitation in question. Among those, the solicited individual considers some of the relevant specific items to be expected to be shared with the interrogator if they were to answer truthfully. For example, if an individual is asked what they did the night before, the activities they performed will be progressively and automatically activated in their memory (i.e., the relevant specific items). However, they expect the person asking to only be interested in them sharing what the interrogated individual sincerely considers to be the main activities they performed worth mentioning (i.e., the relevant specific expected items). The aggregation of all those relevant

specific expected items corresponds to "the" truthful answer to the solicitation. It is important to note that because we consider those expected items to be activated at different times, what is to be considered the truthful answer to a solicitation evolves over time. In other words, the longer an individual has time to think after a solicitation, the more they can remember (up to some limit), and therefore the more complete their version of the truth will be. Also noteworthy is that we consider the truthful answer to be unique, and that only an answer containing *all* the relevant specific expected items activated (and only those) corresponds to the truthful answer, while any answer not containing all the expected items activated in memory at the time of the answer (or any answer containing all the relevant specific items activated *and* other additional items) is deceptive as it at least contains some kind of omission. For example, an individual who is asked what they ate at an event several months ago may not remember (i.e., no relevant specific item activated at this stage, so the truthful answer is "I can't remember"). Five minutes later, however, the person may remember that the starter was soup, and later in the evening they drank some exotic beers they did not like. At this point, a = {soup} is the only relevant specific expected item activated at t = 5 min, and b = {bad-tasting exotic beers} is a relevant specific non-expected activated item at that time. The truthful answer to the question has thus evolved over time from "I don't remember" (time 0 to 5 minutes), to "soup" ($t > 5$ minutes). Put differently, even though their answer has changed, our responder remained sincere and truthful.

Furthermore, besides the relevant items specific to the solicitation, other non-specific relevant items are also automatically activated. Those correspond to bits of memory that are relevant to the solicitation, but do not correspond to activities, opinions, motivations, or intents the interrogated individual truly experienced or believes in. A simple example to illustrate this point is the following: if a thief is interrogated by the police and is asked "did you steal the

money?", the relevant specific item "Yes" would get activated in semantic memory, but probably also the relevant non-specific item "No". In general, we predict that relevant specific items to a solicitation should get activated before the non-specific ones, but that is not necessarily the case (especially if a lie has been rehearsed a lot more than sharing the truth in response to an anticipated question). In any case, we assume the first activations usually occur within only milliseconds or less after a solicitation.

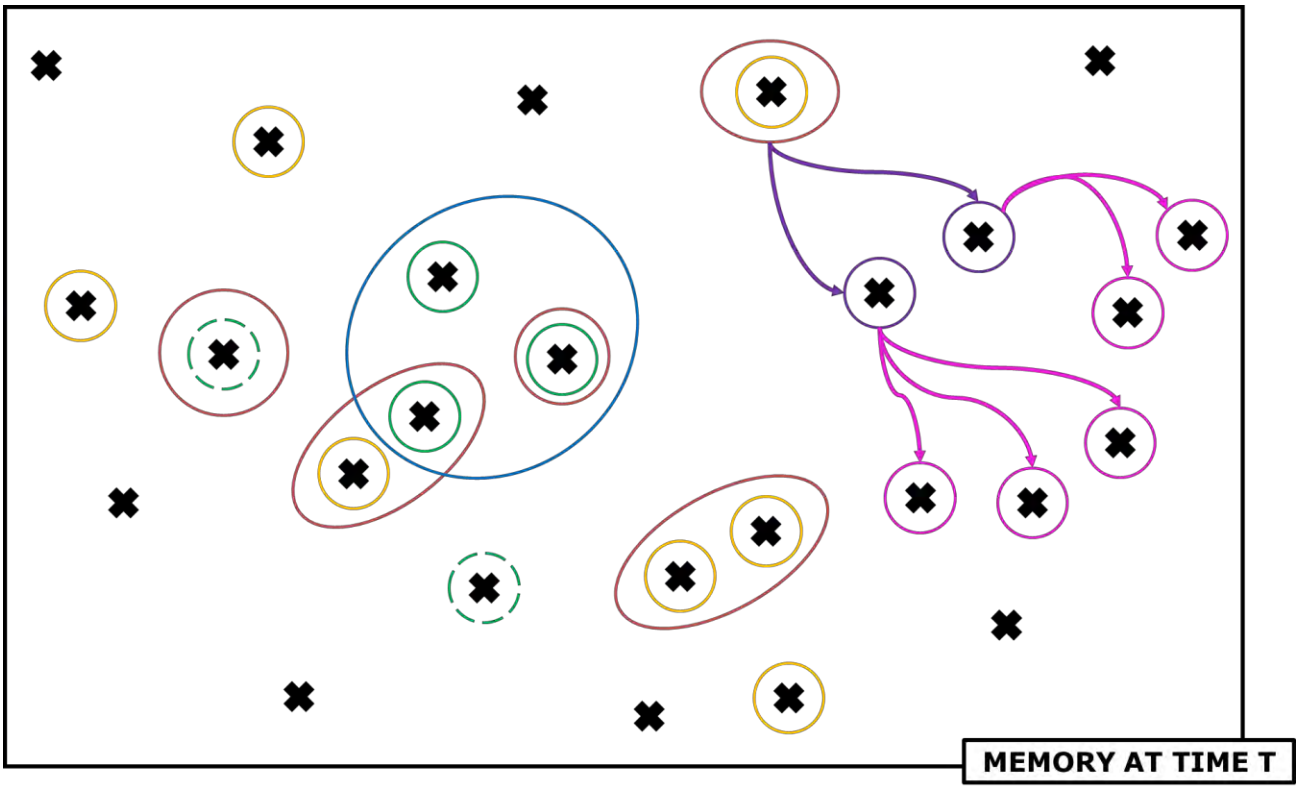
Second, the DDA considers that when a relevant specific expected item gets activated, this automatically triggers an evaluation of whether not disclosing this item ($B_{ND}(item)$) with the soliciting party can have any benefit to the questioned individual. If not, the individual decides (so far) to share the specific item with their counterpart when they answer the solicitation. On the other hand, if they find any potential benefit in not disclosing this specific expected item ($B_{ND}(item) > 0$), then they start to fabricate and evaluate alternatives to sharing the truth. In the DDA, truth is thus considered to be the default state (Levine, 2014). It is only when there is any potential benefit of not disclosing the full truth that people start to consider alternative options to sharing the truth. In other words, the DDA states that people do not even consider lying (or not answering a solicitation) unless they perceive some benefit to it. This should also apply to pathological liars, who could perceive some kind of benefit to deceiving or manipulating others, e.g., by experiencing "duping delight" (Ekman, 2009).

These alternative options to the truth can contain a) only relevant specific items but omit some of the activated relevant specific expected items (lie of omission), b) a mixture of relevant specific and non-specific items (e.g., distortion lies), or c) only a combination of relevant non-specific items (e.g., falsification lies). The evaluations of all the BPRED factors of all the options are automatically launched as soon as an option is activated in memory (once there is at least one

item so that $B_{ND}(item) > 0$). The actual evaluations of the BPRED factors, however, are conscious. The time required to complete a BPRED factor evaluation therefore depends on the complexity of the evaluation, whether similar evaluations have been conducted in the past (i.e., the individual recently faced a similar solicitation for the truth), and the individual's memory capacity at the time. As soon as an alternative to telling the truth is activated in memory, the analysis of all the potential undesirable events that can occur as a consequence of lying with this alternative is also automatically triggered, as well as all the possible justifications the individual could use (i.e., other cover up lies or admitting to the truth) to later justify this lie in case any of the undesirable events they could think of do occur. As soon as an undesirable event is considered in memory, the subjective assessment of the probability of this undesirable event occurring is also automatically triggered, as well as the BPRED factors evaluations of all the justifications considered. Again, all those assessments take different times to complete, depending on the complexity of the task and memory bandwidth at the time.

Figure 1.1

Representation of Memory Activation and Subsets as Modeled in DDA



At some point then, the individual's overall evaluation finishes at time $t = t_{end}$, either because they consider they cannot decently wait longer to answer the solicitation without raising serious doubts ($t = t_{answer}$), or because they have reached the maximum processing time ($t_{processing}$) allowed by their cognitive ability at the time of the evaluation. At this time, the DDA posits that if throughout the process the individual did not find any reason not to disclose a relevant specific item, then they decide to share all the relevant specific expected items (i.e., "the truth") with their counterpart. Otherwise, they choose the option (truth included) with the maximum value/Utility function score. It is worth noting that depending on the activations or BPRED evaluations that had the time to be completed, certain Utility functions could be only partially evaluated at t_{end} and thus lead to a suboptimal decision by the individual. For example, if someone is asked in the street by a stranger "Do you have a cell phone?", they might only have the time to think (a) this person looks odd and might want to steal my phone if I answer truthfully with a "Yes", so (b) I have a benefit in lying. In this case, only the B factor would thus have had the time to be calculated, and the real-time decision to lie would be based only on partial completion of the algorithm.

As noted previously, contrary to the truth option scoring, non-truthful options' scores take into account both the Utility function of the deceptive option itself, and also the Utility functions of their justifications, i.e., the "Justify power" of the deceptive option.

Last, it is logical to assume that the "calculation times" of the DDA algorithm are a function of the following five features: (a) intrinsic cognitive capacities of the individual, (b) contextual cognitive capacities of an individual at a given time (e.g., fatigue limiting available mental resources), (c) number of alternatives and potential DM to process and compare, (d) importance of the Utility functions factors (i.e., the higher the subjective assessments of

the B, P, R, E, D factors, the longer the response time and output generation of the DDA), and (e) proximity of the Utility functions scores between options (i.e., choice becoming more difficult and taking more time when several almost as "good" or as "bad" options are available to the potential deceiver). It could also be questioned whether some DDA factors intrinsically take more time to calculate than others, thus explaining cognitive biases in decision making processes (e.g., the dissonance factor being more quickly evaluated than the cognitive execution factor), but this remains speculative at this stage.

Figure 1.2

Flow Diagram of DDA Consecutives Processes by Which a) Memory Items Are Activated b) Some Evaluations Are Triggered by These Activations, Depending On The Items' Nature c) Each Possible Answering Option Activated is Scored Through a Utility Function, After a Solicitation for Truth.

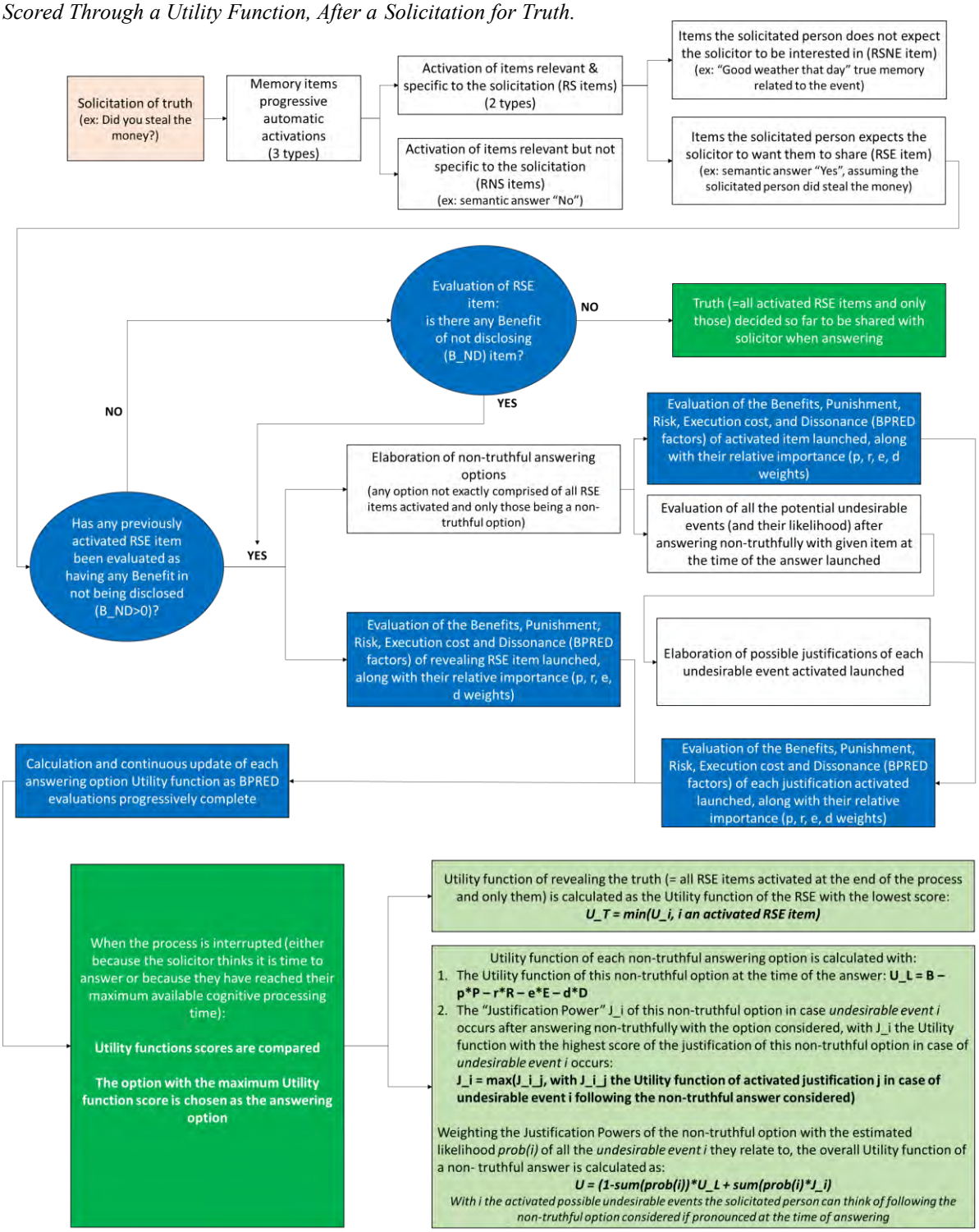


Figure 1.3

Summary of the Conceptual Building Blocks of the GTD.

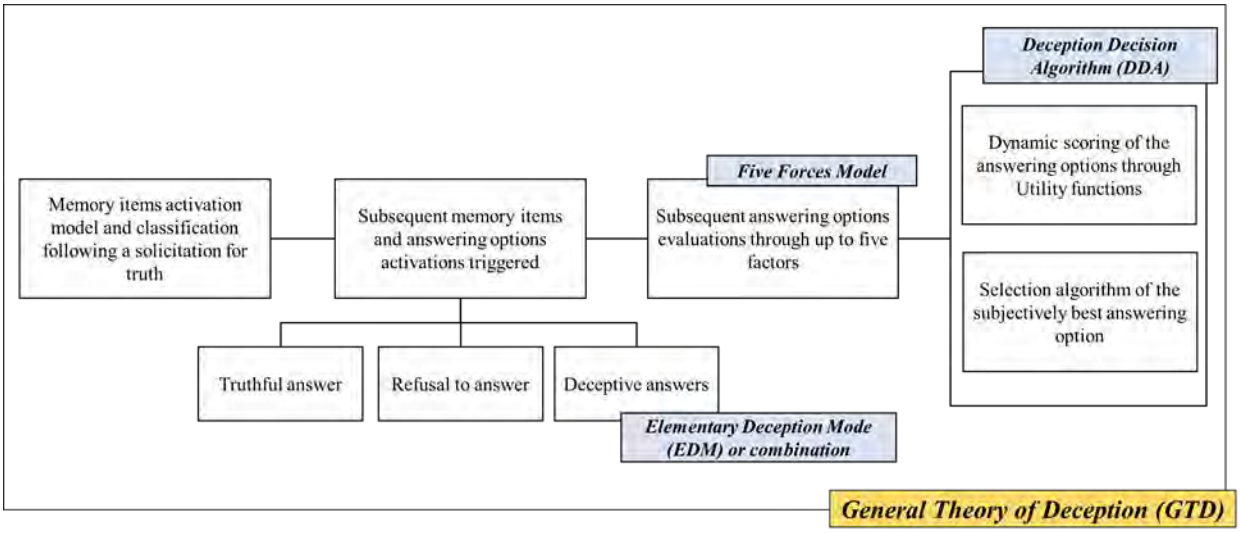


Illustration of the DDA Process: A Simplified Example

Consider the following situation: a 20-year-old youngster used marijuana at a Saturday night party with friends, but did not drink any alcohol. When driving back home around 2am, his car is pulled over by the police. The police officer asks him: "Sir, have you used any drugs or drunk any alcohol tonight?" The youngster considers he has to answer this question quickly (less than 2 seconds) in order to not raise serious suspicions from the police officer.

Applying the DDA, this question by the police officer corresponds to the solicitation of truth, occurring at time $t = 0$. Because the youngster could be thinking of the best answer to provide for more than 2 seconds, it is the timing at which he thinks his answer is expected that is limiting his mental evaluation. Therefore, here, $t_{end} = t_{answer} = 2$ seconds.

As soon as the question is asked to the youngster, memories of the party he attended are automatically and progressively activated: him rolling a joint (*item 1*, a *relevant specific* item in his memory) is activated at time $t_1 = 0.1$ second, the semantic answer “Yes” to the solicitation (*item 2*, a *relevant specific expected* item) is activated at $t_2 = 0.2$ second, the semantic answer

"No" (*item 3, relevant non-specific*) is activated at $t_3 = 0.3$ second, and many other relevant specific items he does not consider the police officer to be expecting in his answer are also activated (e.g., him initially driving to the party activated at $t_4 = 0.4$ second, him drinking a soda right before leaving the party, activated at $t = 0.45$ second, etc.).

In this simplified example, we only considered one relevant specific expected item (*item 2*). Therefore, the DDA is initiated only when this item gets activated in memory at time $t_2 = 0.2$ second. Because this item is a *relevant specific expected* one, the evaluation of the benefits of not disclosing it to the police officer ($B_{ND}(item\ 2)$) is automatically initiated when the item was activated at time $t = t_2$. The evaluation of $B_{ND}(item\ 2)$ completes at $t_5 = 0.5$ second, with the youngster obviously perceiving some benefits of not sharing with the police officer that he used illegal drugs that night (i.e., $B_{ND}(item\ 2) > 0$).

As soon as it is evaluated that $B_{ND}(item\ 2) > 0$ at time $t_5 = 0.5$ second, two processes are automatically triggered in parallel:

1. Evaluation of the BPRED factors and their coefficients of answering truthfully with *item 2* begins. Six-point Likert scales (0 = not important at all, 5 = extremely important) are used to model the importance of each factor of the DDA. The evaluations of the BPRED factors (and their coefficients) respectively complete at $0.6, 0.9, 1.0, 0.8, 0.7$ second, with $B_T\ (item\ 2) = reducing\ mental\ conflict/dissonance\ and\ emotional\ cost\ of\ lying = 1$, $P_T(item\ 2) = getting\ a\ criminal\ record\ and\ a\ fine = 5$, $R_T(item\ 2) = no\ risk\ of\ being\ disbelieved\ if\ admitting\ the\ truth = 0$, $E_T(item\ 2) = high\ emotional\ cost/stress = 4$, $D_T(item\ 2) = no\ mental\ conflict\ of\ admitting\ the\ truth = 0$. The BPRED coefficients are updated each time a new BPRED factor evaluation is completed (see Table 1.2).

2. Elaboration of non-truthful answering options begins. At time t_5, *item* 3 (semantic answer "No", a *relevant non-specific* item) is already activated, and therefore enters the list of possible non-truthful answers to the solicitation (i.e., *item 3* enters the $Activated\_nontruthful(t, solicitation)$ at $t = t_5$) and the evaluation of its BPRED factors and their coefficients also begins at that time. The mental evaluation of the possible undesirable events that can occur if non-truthful *item 3* ("No") is answered by the youngster to the police officer begins as well at $t = t_5$.

Among the evaluations of the BPRED factors of *item 3*, only the evaluations of the B and P factors complete before t_{end}, at respectively $t = 0.65$ and $t = 0.95$ second, with $B_L\,(item\ 3) = avoiding\ criminal\ record = 5$ and $P_L(item\ 3) = criminal\ record + fine = 5$.

At $t = 1.05$ second, when considering lying with *item 3* to the police officer, the youngster thinks of the following undesirable event (*item 4: undesirable event 1 of item 3*): being tested for drugs by the police officer right after answering "No"; an event he estimates has a 50% chance of happening. As soon as this thought of an undesirable event is activated in memory at $t = 1.05$ second, the search for possible justifications he could provide to a positive drug test begins. The youngster will find no justification for his lie in that case before t_{end}.

At $t = 1.2$ second, a new non-truthful option is activated (*item 5*), consisting of telling the police officer: "Officer, I haven't drunk a sip of alcohol!" (*EDM 8 – Selective Response*). This thought is per se a *relevant specific* item, but because this answering option does not mention *relevant specific expected item 2*, it represents a lie of omission. As soon as this thought (*item 5*) was activated, its BPRED factors evaluation started (which for the sake of simplicity we assume both completed for the $B_L\,(item\ 5)$ and $P_L\,(item\ 5)$ factors at $t = 1.25$ second and were all equal to those of *item 3*, the remaining factors evaluations not finishing before t_{end}), as well as

the evaluation of the possible undesirable events that can occur afterwards. Again, the drug test undesirable event (50% probability) is quickly considered at $t = 1.3$ second in that case too (*item 6*), as it was already activated in memory relative to *item 3*. But with that deceptive answering option the youngster comes to think, at $t = 1.6$ second, that at least he could say, if tested positive, he did not lie per se as he only said he did not drink alcohol; he did not deny using drugs (*item 7*). As soon as this justification was activated in memory, its BPRED factors evaluation was launched, but only the evaluation of the factor $B_L(item\ 7) = reducing\ mental\ conflict/dissonance\ by\ not\ technically\ lying = 3$ completes before t_{end}, at $t = 1.9$ second.

At $t = t_{end} = 2$ seconds, while some mental evaluations were still on-going, the (partial) Utility functions scores of the different options the youngster had to time to calculate are compared, with $U_T(item\ 2) = -1.25$, $U_L(item\ 3) = 0$ and $U_L(item\ 5) = 1.5$. Because lying with *item 5* ("I haven't drunk a sip of alcohol") has the highest score, this option is selected by the youngster as his best answer to the police officer.

Table 1.2

Representation of the Mental Processes and DDA in the Police Control Example

Time stamp	Time (sec)	Event occurring	Type of item	Evaluation result	Utility function score	Consecutive events triggered
t_0	0.00	Solicitation of truth (have you used any drugs or drunk any alcohol tonight)	Solicitation	-	-	Activation of relevant items in Memory begins
t_1	0.10	Activation of item 1 (rolling a joint)	Relevant specific	-	-	-
t_2	0.20	Activation of item 2 (semantic answer Yes)	Relevant specific expected	-	-	Evaluation of $B_{ND}(item\ 2)$ begins
t_3	0.30	Activation of item 3 (semantic answer No)	Relevant non-specific	-	-	-
t_4	0.40	Activation of item 4 (driving to the party)	Relevant specific	-	-	-
t_5	0.50	Evaluation of $B_{ND}(item\ 2)$ completes	B_{ND}	$B_{ND}(item\ 2) > 0$	-	• Evaluation of item 2 BPRED factors and weights begins • Elaboration of non-truthful answering options begins. Item 3 (already activated in Memory) immediately enters Activated_nontruthful(t_5, solicitation), and the evaluation of all the potential undesirable

Time stamp	Time (sec)	Event occurring	Type of item	Evaluation result	Utility function score	Consecutive events triggered
						events that can happen after t_{end} if lying with item 3 at t_{end} begins • Evaluation of item 3 BPRED factors and weights begins
t_6	0.60	Evaluation of $B_T(item\ 2)$ completes. Utility function of item 2 $U_T(item\ 2)$ is calculated	B	$B_T(item\ 2) = 1$	$U_T(item\ 2) = 1$	-
t_7	0.65	Evaluation of $B_L(item\ 3)$ completes. Utility function of item 3 $U_L(item\ 3)$ is calculated	B	$B_L(item\ 3) = 5$	$U_L(item\ 3) = 5$	-
t_8	0.70	Evaluation of $D_T(item\ 2)$ and $d_T(item\ 2)$ complete. $U_T(item\ 2)$ is updated	D, d	$D_T(item\ 2) = 0$ $d_T(item\ 2) = 1$	$U_T(item\ 2) = 1 - 1 * 0 = 1$	-
t_9	0.80	Evaluation of $E_T(item\ 2)$ and $e_T(item\ 2)$ and $d_T(item\ 2)$ reevaluation complete. $U_T(item\ 2)$ is updated	E, e, d	$E_T(item\ 2) = 4$ $e_T(item\ 2) = \frac{1}{2}$ $d_T(item\ 2) = \frac{1}{2}$	$U_T(item\ 2) = 1 - \frac{1}{2} * 4$ $- \frac{1}{2} * 0 = -1$	-

Time stamp	Time (sec)	Event occurring	Type of item	Evaluation result	Utility function score	Consecutive events triggered
t_{10}	0.90	Evaluation of $P_T(item\ 2)$ and $p_T(item\ 2)$ and $e_T(item\ 2)$ / $d_T(item\ 2)$ reevaluation complete. $U_T(item\ 2)$ is updated	P, p, e, d	$P_T(item\ 2) = 5$ $p_T(item\ 2) = \frac{1}{3}$ $e_T(item\ 2) = \frac{1}{3}$ $d_T(item\ 2) = \frac{1}{3}$	$U_T(item\ 2) = 1 - \frac{1}{3} * 5 - \frac{1}{3} * 4 - \frac{1}{3} * 0 = -2$	-
t_{11}	0.95	Evaluation of $P_L(item\ 3)$ and $p_L(item\ 3)$ complete. $U_L(item\ 3)$ is updated	P, p	$P_L(item\ 3) = 5$ $p_L(item\ 3) = 1$	$U_L(item\ 3) = 5 - 1 * 5 = 0$	-
t_{12}	1.00	Evaluation of $R_T(item\ 2)$ and $r_T(item\ 2)$ and $p_T(item\ 2)$ / $e_T(item\ 2)$ / $d_T(item\ 2)$ reevaluation complete. $U_T(item\ 2)$ is updated	R, p, r, e, d	$R_T(item\ 2) = 0$ $p_T(item\ 2) = \frac{1}{4}$ $r_T(item\ 2) = \frac{1}{4}$ $e_T(item\ 2) = \frac{1}{4}$ $d_T(item\ 2) = \frac{1}{4}$	$U_T(item\ 2) = 1 - \frac{1}{4} * 5 - \frac{1}{4} * 0 - \frac{1}{4} * 4 - \frac{1}{4} * 0 = -1.25$	-
t_{13}	1.05	Activation of item 4 "undesirable event 1 of item 3" (being tested for drugs) and associated probability evaluation completion. Item 4 enters Undesirables_item3(t_{13})	Undesirable event	$prob\ (undesirable\ event\ 1\ of\ item\ 3) = 50\%$	-	Elaboration of possible justifications of item 4 begins

Time stamp	Time (sec)	Event occurring	Type of item	Evaluation result	Utility function score	Consecutive events triggered
t_{14}	1.20	Activation of item 5 (I haven't drunk a sip of alcohol)	Relevant specific	-	-	• Item 5 immediately enters Activated_nontruthful(t_{14}, solicitation) • The evaluation of all the potential undesirable events that can happen after t_{end} if lying with item 5 at t_{end} begins • Evaluation of item 5 BPRED factors and weights begins
t_{15}	1.25	Evaluation of $B_L(item\ 5)$, $P_L(item\ 5)$ and $p_L(item\ 5)$ complete. Utility function of item 2 $U_L(item\ 5)$ is calculated	B, P, p	$B_L(item\ 5) = 5$ $P_L(item\ 5) = 5$ $p_L(item\ 5) = 1$	$U_L(item\ 5)$ $= 5 - 1 * 5$ $= 0$	
t_{16}	1.30	Activation of item 6 "undesirable event 1 of item 5" (being tested for drugs) and associated probability evaluation completion. Item 4 enters Undesirables_item5(t_{16})	Undesirable event	$prob$ $(undesirable\ event\ 1\ of$ $item\ 5) = 50\%$	-	Elaboration of possible justifications of item 6 begins
t_{17}	1.60	Activation of item 7 (I only said I did not drink alcohol, not not used drugs). Item 7 enters	Justification	-	-	Evaluation of item 7 BPRED factors and weights begins

Time stamp	Time (sec)	Event occurring	Type of item	Evaluation result	Utility function score	Consecutive events triggered
		Justifications_Undesirable1_of_item5(t_{17})				
t_{18}	1.90	Evaluation of $B_L(item\ 7)$ completes. JustificationPower_item7_item6_item5 is calculated and Utility function of item 5 $U_L(item\ 5)$ is updated	B	$B_L(item\ 7) = 3$ =JustificationPower_item7_item6_item5	$U_L(item\ 5) = (1 - 50\%) * (5 - 1 * 5) + 50\% * (3) = 1.5$	-
t_{end}	2.00	Time to answer is reached. Utility scores of the different answering options are compared and the option with the highest value is selected as the best option	Decision-making	max($U_T(item\ 2)$, $U_L(item\ 3)$, $U_L(item\ 5)$) =$U_L(item\ 5)$	-	The youngster answers the police officer with deceptive item 5: "Officer, I haven't drunk a sip of alcohol!"

Notes:

(1) Not all relevant specific items activated before t_{end} are listed in Table 1.2.

(2) To simplify the example, all the p, r, e, d weights were assumed to be equal.

Discussion

The General Theory of Deception (GTD) is a new theory of lie production comprising three distinct parts. The EDM (and their combinations) describe all the options people have at their disposal to deceive their counterpart. When an individual is solicited for the truth, up to five factors are mentally evaluated to decide whether to tell the truth or lie (and if so, with which deception mode). The Five Forces Model describe those five factors (Benefits, Punishment, Risk, Execution, Dissonance), and the DDA describes the process by which their evaluations are mentally and dynamically performed. It is disruptive in the sense that it is the first deception theory to have modeled, from end-to-end, the mental processes involved from the solicitation for truth to the decision to disclose the truth or lie, how truth and lies are activated and elaborated in memory, and how those different options are dynamically evaluated in the individual's mind. Most importantly, for the first time in the deception literature, this decision-making process is described through a specific algorithm and not with general principles, making the theory testable experimentally but also allowing for new predictions.

First, special cases often studied in the deception literature can be well described and explained within the framework of the GTD with the Five Forces Model, which states that five (and only five) BPRED factors (Benefits, Punishment, Risk, Execution cost, Dissonance) are weighted against each other to evaluate whether to lie or tell the truth. Confessions, for example, correspond to cases where the Justification power of the deceptive options considered is very low, or the burden of lying over time has made the benefit of admitting the truth, thus clearing one's conscience, very high. One the other hand, false confessions (e.g., admitting to a murder one has not actually committed) correspond to situations where telling the truth is perceived as an option with a higher risk of being disbelieved (e.g., a police officer telling the suspect that

they know the suspect is guilty anyway) and a bigger punishment ("just admit you did it and I will make sure your punishment is more lenient) than lying by making a false confession. Likewise, individuals persisting in their lies even when confronted with incriminating evidence correspond to situations where individuals perceive more benefits in preserving their image than costs to lying.

Social white lies can also be well described with the Five Forces Model. The current prevailing research paradigm assumes that lying is more cognitively demanding than truth-telling, for several reasons (Vrij et al., 2011). While such statements might certainly be true for many real-life high-stakes lies, like criminal cases, it may not apply to more common everyday social white lies. In the GTD, we do not assume that lying necessarily has a higher execution cost (either cognitive, emotional, or control/inhibition cost) than telling the truth. In social lies, for example, when receiving a sweatshirt gift you do not like at all from your dear aunty Alicia you sincerely love, it is likely that you would find it just more simple (lower Execution cost) to tell her you like the gift, smile, and thank her, rather than explaining her you thank her for the gift but that you do not like it and would never wear it.

Crucial to the elaboration of the GTD was the initial attempt to identify all the ways people behaved in real life to convey deceptive messages, which we coined the "Elementary Deception Modes" (EDM). As a result of a 3-year observational (and introspection) work, 98 verbal and nonverbal EDM have been identified, grouped into categories (Verbal vs. Nonverbal, Deceptive Omission, Falsification, Distortion, Misinterpretation, Deceptive Evasion, Deceptive Diversion), named and exemplified – an analysis never performed before in the literature to our knowledge. As EDM can obviously be combined during a deceptive episode, the billions of possible combinations ($2^{98} - 1$ exactly, or approximately 10^{29} combinations) give a sense of the many

subtle ways people can deceive their counterparts in real life – an important difference to the BFL only paradigm used in experimental deception research to date. Because the GTD describes the mental decision-making process used by individuals through the rules of the DDA algorithm, it can also predict which EDM should be more or less prevalent and frequent in real life, and modulated by the context of solicitation for the truth or an individual's characteristics. For example, in a given context, *EDM 1 – Narrative Omission* is more likely to be frequently used than *EDM 14 – Narrative Fabrication*, as it requires less cognitive effort (lower E factor), is considered more socially acceptable (lower D factor), and is easier to justify if later questioned (higher Justification power J, arguing one has not lied per se). Predictions can even be made for non-falsification EDM. For example, a wife who had an infidelity with a stranger while traveling for work and is questioned by her suspicious husband about her whereabouts on the night of the infidelity, could choose to describe everything she did that night, except for the infidelity (*Deceptive Omission EDM 1 – Narrative Omission*). Alternatively, she could change the subject, saying something like, "Look, the journey here was very long, and then the hotel was really disappointing, so this trip did not really start well!" (*Deceptive Diversion EDM 77 – Topic Switch*). In GTD terms, the wife is more likely to opt for the former, because she might fear that changing the subject might raise suspicions from her husband ($R^{EDM\ 77} > R^{EDM\ 1}$) and lead to more follow-up questions.

Our observation of the EDM and the huge number of their possible combinations also tends to "lower the bar" of what people would spontaneously consider lying; and question previous findings that the prevalence of deception reported by participants in real life is low and its distribution skewed with only a few prolific liars (Serota et al., 2021; Serota & Levine, 2015). Much of communication is subtle, with meaning imparted implicitly. It might be argued, for

example, that some of the evasion and diversion strategies identified in our 98 EDM-list are not actual forms of deception but rather, implicit messages that one does not wish to answer a question. So how does the GTD differentiate the socially skillful act of changing the subject, or a signal that one will not answer a question, from deception? According to the definition of deception we use (Vrij, 2008), it is the fact that the person that is questioned has the deliberate intent, or not, to create a wrong belief in the questioner that will make any evasion or diversion a lie. To illustrate this, let us take a concrete example (the one we used to describe *Deceptive Diversion EDM 79 – Exasperation card*). A wife asks her husband: “Who are you having dinner with tonight?”. The husband answers: "Oh come on stop bothering me with all your questions every time I have a dinner!". Is this a socially skillful act of changing the subject? Or a deliberate intent to deceive? Our view is that it all depends on the intent of the husband: if he has nothing to hide and is just truly exasperated by an overly suspicious spouse, then this diversion answer (it is a diversion as he did not answer the question asked) is not a lie. On the other hand, if the husband is having dinner with his mistress, then this diversion answer is meant to deceive his wife by making her believe he is exasperated with her undue jealousy, while really he just wants to avoid fabricating a story about whom he is going to have dinner with. In other words, an answer can be deceitful or not depending on the intent to create a false belief in the solicitor, or not. More specifically, how does the GTD differentiate between the following scenarios:

(a) a husband that has nothing to hide from his wife but was truly exasperated (and still did not answer his wife’s question about whom he was having dinner with);

(b) a husband having nothing to hide, being exasperated, but answering the question, e.g., “If you must know, I am having dinner tonight with Peter [his boss] to discuss my quarterly

objectives. But I would like you to stop bothering me with all your questions every time I'm out for dinner!"; and

(c) a husband having dinner with his mistress and deceiving by feigning exasperation?

In the GTD's Deception Decision Algorithm, this would mean that the answering option with the highest Utility function value is:

- For (a): U_{NA}, the Utility function of not answering a question (which is different from both telling the truth and lying);
- For (b): U_T, i.e., the option of answering truthfully;
- For (c): U_L^{79}, the Utility function of answering deceitfully with the lie option *Deceptive Diversion EDM 79 – Exasperation Card*.

Much in the same vein, how does the GTD address exchanges mostly dictated by social norms, for example when someone is simply asked "How are you?". Arguably, it would be considered a communication error to specifically enumerate all the factors that are having an impact on the responder's state of well-being at that particular moment. Would someone answering "Great" in a routinized manner, as is often the case, be lying? In the GTD's modeling, there is only one truthful answer, at a given time, to a solicitation: the one containing all the *relevant specific expected* items activated in memory at time t (and only those). In our social greetings example, some relevant specific items activated after the individual is asked "How are you?" could include a = {Great} and b = {I've had a slight headache for the last 10 minutes}, while relevant non-specific item c = {Not so good} could also be activated (with (c) being semantically relevant but not corresponding to the current state of the responder). As the responder believes only (a) is expected in response to this question (i.e., (a) is the only relevant specific expected item activated), as dictated by social conventions, then (a) is "the" truthful

answer in this example – even though (c) was not shared with the solicitor. Now, if the solicitor follows this up with "Are you sure? Something seems to be bothering you", then (c) would probably become a relevant specific expected item as well. The context, questions asked, and overall communicative interplay dictate the dynamic expectations over time and assessment of what the truth includes or not, in the mind of the person questioned. This subjective evaluation of the solicitor's expectations requires Theory of Mind, i.e., the ability to imagine the thoughts, feelings, knowledge, or intentions of others (DePaulo, 1992; Gombos, 2006; Talwar et al., 2007). However, this does not imply that populations with lower Theory of Mind abilities are not as able to provide relevant expected answers; rather, what they believe to be a relevant expected answer could be more likely to be perceived as irrelevant by the solicitor.

Knowledge of the EDM and the GTD's DDA have practical implications for rethinking and identifying interviewing practices that can enhance the interviewer's ability to detect lies. For example, the current literature on the best interviewing practices to detect deception generally recommends open-ended questions (e.g., *"Tell me everything you did the day of the theft"*), as opposed to close-ended direct questions (e.g., *"Did you steal the money from the envelope?"*). This recommendation works on the assumption that the latter requires less cognitive effort (i.e., a low E factor in the GTD) from liars to replace a true "Yes" with a deceptive "No" answer (*EDM 15 – False Response*), as opposed to having to fabricate the narrative of activities as an alibi (*EDM 14 – False Narrative*). We, on the contrary, believe such closed-ended questions can be very useful after an information-gathering part of an interview. More specifically, if in GTD terms such direct questions indeed result in lower E factors compared to open-ended questions ($E^{EDM\ 15} \ll E^{EDM\ 14}$), they on the other hand create a much higher mental conflict ($D^{EDM\ 15} \gg D^{EDM\ 14}$) and are also much harder to justify if later caught lying ($J^{EDM\ 15} \ll J^{EDM\ 14}$), as direct

close-ended questions do not allow any room for ambiguity and thus neither do direct false responses. Thus, defining the exact suspicion of deception one has at the beginning of an interview (and updating it during the course of the interview given the interviewee's answers) is of paramount importance to properly use direct questions. However, this is not always possible; for example one cannot ask a car seller if a specific part of the vehicle has to be replaced soon if they do not even suspect that kind of scam from the seller.

Additionally, the meticulous work performed to identify EDM also provides insights on how to better frame questions and evaluate answers for potential deception during interviews. Consider the example provided to illustrate DDA with *EDM 8 – Selective Response*: a 20-year-old youngster used marijuana at a Saturday night party with friends, but did not drink any alcohol. When driving back home around 2am, his car is pulled over by the police, who asks him: "have you used any drugs or drunk alcohol tonight?" The youngster quickly answers: "Officer, I haven't drunk a sip of alcohol!" Because this EDM relies on the interrogator asking multiple questions within a single solicitation sentence, a straightforward way for an interrogator to avoid an interviewee making use of such a DM is to separate the questions (e.g., *Question 1: "Did you use drugs tonight?", Question 2: "Did you drink alcohol tonight?"*), and thus forces the interviewee into more costly falsification lies as their only deceptive options. In the simulation of this case described in the present paper (Table 1.2), we thus see that removing the *Selective response* deceptive option (with a Utility score $U_L = 1.5$), by separating the police officer's questions, would have led the youngster to have to opt for a less advantageous deceptive answer (*EDM 15 – False Response*), of Utility function score $U_L = 0$.

The coding of potential EDM in a statement transcript can, in and of itself, be used as a credibility assessment tool – the more potential EDM coded the more likely the statement will

contain deception. Yet, the nuance provided by the identification of the exact potential EDM used in a statement can prove impractical for professionals, given the extensive list of 98 EDM identified in our work. To reduce this complexity, we propose to introduce one basic (and therefore, applicable in practice) deception detection principle derived from the GTD. *Nobody Wants to Lie*. In the GTD's Deception Decision Algorithm, the different options one mentally compares in response to a solicitation for truth are modeled using Utility functions. The option with the highest score is the one selected for answering the solicitation. Between the deceptive options mentally considered, we posit that of all the deception modes (from the EDM categories defined in the GTD, i.e., Deceptive Omission, Falsification, Misinterpretation, Distortion, Deceptive Evasion, and Deceptive Diversion), those the furthest from Falsification (such as Deceptive Evasion and Deceptive Diversion), will generally have the highest Utility functions scores. While this statement of course depends on the context and the exact way the questions are framed to the interviewee, most of the time, non-Falsification deception modes should be preferred as they lower the Execution cost factor (it is less cognitively and emotionally demanding to omit a detail than to fabricate one, for example), the Dissonance factor (falsifying an answer is more mentally and morally conflicting than dodging a question), and the Punishment factor (people believe they will have less severe punishments if they think they "did not really lie", as could be the case with non-Falsification deception modes), and for the same reason, increase the perceived Justification power of such deception modes ("I can always say I did not really lie"). For these reasons, the GTD predicts that the Five Forces push people to opt for a deception that is the closest to what they believe is the truth. Consider now the statement analysis principle that "Nobody wants to lie". This does not mean we believe that nobody lies. Rather, this means that people, when they do not want to disclose the truth, try as much as

possible to be deceptive while keeping the conscious or unconscious impression they are not really lying. This should tend to make them minimize the deviation of their answers to the truth. This principle can indeed, we believe, be a very simple yet efficient way to better detect deception, especially in response to direct questions addressing a clearly defined suspicion, which we illustrate below.

As a simple example (the one chosen to illustrate *EDM 68 – Implied Response*), imagine a police officer asking a suspect: "Did you kill your wife?". The latter responds: "I have never hurt anybody!". Recognizing that the police officer's suspicion is whether the suspect killed his wife, and applying the *Nobody wants to lie* principle, the police officer should immediately notice that the suspect could actually have killed his wife (the suspicion) while thinking he did not make her suffer in the process, making the suspect (unconsciously?) believe he technically did not lie, and probably reducing his Dissonance factor with such an answer. Using this principle with a clearly defined suspicion can thus, we believe, allow practitioners to detect traces of all non-Falsification deception. Combining the *Nobody wants to lie* principle with direct final questions addressing the suspicion at stake thus forces the deceptive interviewee to choose one of three options:

(a) reiterating non-Falsification lies, which as described above leave direct easily detectable traces (i.e., repeatedly not answering a specific direct question with a perfectly specific answer);

(b) opting for a Falsification lie;

(c) getting away from the previous ambiguity of non-Falsification lies, and telling the truth.

For example, going back this time to the example used to illustrate *EDM 1 – Narrative Omission*, imagine a husband told his wife he would be home from work early today, but ends up arriving more than an hour and a half late. When his wife asks him what happened, he answers:

"I had a lot of work today". Even though he indeed had an intense day at work, he fails to mention he still found some time to have a drink with his colleagues after work before heading back home. If, in similar situations in the past, his wife had discovered that he had "wasted" some time with colleagues before heading back home, she could now have the same suspicion whenever he comes home late. She therefore asks him this time: "Did you come straight home after work?" (a direct question addressing the suspicion). The husband is now forced to either persist in non-Falsification deception ("Honey, you know how hard I have been working lately, let's not waste any more time and start our evening"), admit the truth ("No, I had some beers with colleagues after work first") or engage in a Falsification lie ("Yes, I came straight home after work").

Following the interviewing method we suggest should thus lead to either easily detectable lies, people telling the truth, or falsification (and all categories of nonverbal) lies that leave no direct traces detectable with the *Nobody wants to lie* principle. What to do in such cases? Fortunately, as developed above, the GTD predicts that those lies should be the less prevalent ones, and yet they can also be the most important to detect. To address this specificity, we propose that credibility assessment interviews include what could be called "External and Internal Consistency Tests". Simply put, the interviewer should in such cases orientate and asks questions whose answers they can compare against (a) external facts or evidence (External Consistency Test), or (b) other previous answers from the interviewee asked in different formats (Internal Consistency Test). Here is an example of an External Consistency Test. If someone claims to have finished a regional marathon in 10th place, the interviewer should orientate their questions to get a maximum of verifiable information about the event, and then actually check (if any) the published marathon ranking on the internet. Although both asking for verifiable details

and checking whether the interviewee has provided verifiable details in their statement have already been suggested and successfully tested as a cue of truth telling in the literature (Harvey et al., 2017; Jupe et al., 2017; Nahari & Vrij, 2014; Vrij et al., 2016), here, we go a step further. Simply put, we state that *what is better than a verifiable detail, is a verified detail*. In other words, we recommend not only checking whether an interviewee has provided verifiable details, but trying as much as possible to orientate the interview questions to have interviewees state information that can be verified, try to verify it, and confirm or contradict the interviewee's statement. Then, if the nature or context of a statement justifies the absence of external corroborating evidence (e.g., someone talking about an incident that happened at work during a confidential business meeting), the interviewer should resort to Internal Consistency Tests. For example, the interviewer could first ask the interviewee to verbally describe, in as much detail as possible, who was seated where around the negotiation table, and much later in the interview, to sketch the configuration and protagonists around the table, then compare the consistency of the answers. Conducting such tests during the interview, especially External Consistency Tests, also have the benefit, in GTD terms, of increasing the deceitful interviewee's perception of "the likelihood of an undesirable event" of lying (i.e., that their version will be checked and might be contradicted by external evidence), thus decreasing the Utility functions scores of their deceptive options.

Most importantly, the main practical prediction and implication of the GTD is that the understanding of the mechanisms described by the DDA and the Five Forces Model can allow an interviewer to manipulate a context and the framing of their questions to make the interviewee more honest. More specifically, because the GTD models that an individual bases their decision to lie or tell the truth on their subjective assessment of the BPRED factors of all the potential

answers they can think of in response to a solicitation, a direct consequence is that *artificially* changing the individual's perception of the BPRED factors could not only make cues to deception more salient, but could even prevent the individual from engaging in deception in the first place. While the current literature only recently focused on increasing cognitive load (E factor in GTD) to make the cues to deception more diagnostic (e.g., by asking unexpected questions) or to deter people from engaging in deception (Van 't Veer et al., 2014), the GTD predicts more largely that the likelihood that an individual will engage in deception can be manipulated by any combination of the following: increasing the perceived benefits of telling the truth, decreasing the perceived benefits of lying, decreasing the PRED factors of telling the truth, and/or increasing the PRED factors of lying.

For example, if a little boy promised his mother he would go straight back home after school and ends up playing football with his friends before going home (*Situation 1*), he would face a greater mental conflict when lying to his mother once home than if he had lied in the exact same way about his activities but without having previously made this promise (*Situation 2*). In GTD terms, we would say that the dissonance $D_L^i(Situation\ 1)$ caused by lying with a given deception mode i in *Situation 1* is higher for the boy than if he had not previously promised anything to his mother ($D_L^i(Situation\ 1) > D_L^i(Situation\ 2)$), thus diminishing (but not necessarily preventing) the likelihood of the boy engaging in deception. This "freezing effect" of past decisions and commitments is a well-documented phenomenon (Bazerman et al., 1984; Drummond, 1994; Geiger et al., 1998; Lewin, 1947; Moon, 2001a, 2001b; Moriarty, 1975; Ross & Staw, 1993; Staw & Ross, 1989), and the GTD posits that the same effect can be observed by manipulating all the BPRED factors. It is noteworthy however that the GTD does not assume BPRED factors to be totally independent and uncorrelated, i.e., manipulating and increasing one

factor can also lead to decreasing another one by the same process. For example, if a deceptive answering option (2) decreases the mental conflict and dissonance D felt by the solicited person compared to an alternative deceptive option (1) they were considering (in other words, $D(option\ 2) < D(option\ 1)$), it is likely that the emotional component of the execution cost E of option (2) will also be lower than the emotional component of the execution cost E of option (1) (in other words, $E(option\ 2) < E(option\ 1)$).

Chapter 2: Deception in Negotiations: Making people more honest with a two-factor approach

Dematerialized negotiations are increasing and tend to foster deceptive exchanges, while also limiting the ability to detect signs of deception. We therefore tested whether it was possible to deter negotiators from producing deceptive statements in the first place, by manipulating the mental conflict of lying (Dissonance factor) and increasing the perceived risk of being disbelieved if lying (Risk factor). A total of 458 participants were recruited online and placed in a buyer/supplier email negotiation scenario in one of eight experimental groups, manipulating the Dissonance factor (using, or not, a "you have the reputation of being honest and trustworthy" pro-social labeling technique), and Risk factor (three price levels of an alleged alternative offer and a control group). They were ultimately, as suppliers, asked their own purchasing price by the buyer – a highly sensitive piece of information. Results showed that the proportion of participants giving a true price in response to the question of their own purchasing price was significantly higher when they were exposed to our deception deterrence factors. Of the participants who did not dodge the question entirely, 38.1% of the control condition (no Dissonance, no Risk factor) gave their true purchasing price, vs. 80.8% in the combined Dissonance/Highest Risk factor level condition. Professionals can thus use simple theory-based deception deterrence factors to make their counterpart more honest in negotiations.

Introduction

Deception research has traditionally focused on trying to identify cues to deceit, mostly in forensic contexts, with mitigated results regarding the discriminative power of individual verbal or nonverbal behaviors (DePaulo et al., 2003). In typical deception experiments, participants either commit a mock crime (e.g., stealing money), or perform normal activities. An interrogator then asks participants open-ended questions, such as "tell me everything you did for the last 30 minutes", to determine whether they are lying when they state they did not commit the mock crime. Such research paradigms attempt to replicate criminal-like interview settings. They lead to the collection of open-ended statements from participants, required to apply well-documented credibility assessment tools that rely on sufficiently long narratives. Examples of such tools include Criteria Based Content Analysis (Köhnken and Steller, 1988; Steller and Köhnken, 1989) and Reality Monitoring (Johnson and Raye, 1981). In most real-life situations, however, collecting open-ended narratives in order to apply the traditional credibility assessments tools is simply not possible.

While there are some findings related to the detection (Van Swol et al., 2012; Belot and van de Ven, 2017) and prevention (Aquino, 1998; Olekalns and Smith, 2007; Stawiski et al., 2009; Olekalns et al., 2014) of deception during business negotiations, these are relatively few. Furthermore, they are based on findings with students rather than professionals, and mainly focused on creating "external" reasons not to lie (e.g., by creating an "ethical climate"), rather than manipulating internal more engaging reasons to do so. This is a significant gap in the literature because, for business, applied studies investigating deception during negotiations can produce real financial benefits. Thus, the aim of this study is to test two easily implementable techniques as a means of reducing deception in a business negotiation scenario.

Deception Production Theories

Deception production theories have only been produced very recently (McCornack et al., 2014; Walczyk et al., 2014); thus they are relatively new and scarce within the literature. Information Manipulation Theory 2 (IMT2; McCornack et al., 2014) is a theory of deceptive discourse production, developed as a refinement of the original Information Manipulation Theory (IMT; McCornack, 1992). The central premise of IMT2 is that deceptive and truthful discourse both emerge from one speech production system, not from distinct production systems. This can explain why cues to deceit are subtle (DePaulo et al., 2003) and often only observable in the very specific contexts of 100% fabrication, referred to by the author as bald-faced lies. Second, IMT2 contrasts the prevailing literature's argument that deceptive speech begins with an initial conscious decision to lie (McCornack et al., 2014). Instead, it argues that deception is performed incrementally during speech, as an attempt to provide the best answer to a problematic prompt. This problem-solving task is dynamically guided by accessing the most easily available information in memory. A consequence of this incremental problem-solving process is that people will commonly alternate between truth and deceit during the course of a speech, as opposed to mentally creating fully deceptive messages before articulating them.

The Activation-Decision-Construction-Action Theory (ADCAT; Walczyk et al., 2014) acknowledges IMT2's explanation of spontaneous lying within normal conversation, but claims it is less applicable to high-stakes lies. ADCAT postulates that the choice of whether and how to deceive is based on the expected value of a decision ("EV"), which is calculated by weighing up the probability of a particular outcome, and multiplying it by the desirability of that outcome. ADCAT posits that when truth is solicited in high-stakes situations, respondents compare the EV of their different options (e.g., tell the truth, understate the truth, totally falsify the truth, etc.) and

choose the option with the highest EV. It could therefore be argued that this is the cognitive mechanism behind deception during business negotiations, as these are often high-stakes contexts.

Deception in Negotiations

In bargaining contexts, negotiators often interact dynamically and in a bilateral fashion, seeking to uncover important specific information. For example, a negotiator may try to determine what minimum price offer their counterpart would deem acceptable to assess whether a zone of possible agreement between the negotiators is possible (Fisher et al., 2011). On the other hand, a negotiator making a full and honest disclosure of such strategic information could struggle to maximize their short-term profits (Boles et al., 2000). Accordingly, negotiators may mentally calculate the EV of telling the truth compared to providing misleading answers or omitting strategic information during the bargaining process. A car buyer may claim to have a better offer from another dealer, or the seller may falsely claim they would lose money if selling below a given floor price. Negotiation is fraught with deception and, in a capitalist economy, society often sees it as acceptable or at least expected (Carr, 1968; Strudler, 1995), even when such approaches can compromise trust and longer-term collaborations (Peters, 1987). As illustrated by Strudler (1995: 805):

> The truth can get in the way of a good deal. So, many people lie, dissimulate, and otherwise fail to tell the truth in negotiation … I will maintain that despite the common-sense moral presumption against deception more generally, some deception in negotiations, including lies about one's reservation price, may be morally acceptable.

In the business world, such deception, often referred to as "bluffing", can include making factually inaccurate statements, omitting strategic information, dodging sensitive questions, or exaggerating/minimizing some information (Aquino & Becker, 2005).

A comparison of deceptive and truthful monetary exchanges revealed that individuals who were lying stated their offer more times and gave more supporting statements for their offer (Van Swol et al., 2012). Interestingly, their counterparts did not perceive these cues, and were unable to detect deception above the level of chance. Conversely, Belot and van de Ven (2017) found that participants were able to detect deception above the level of chance in an artificial selling scenario. However, the underlying cues to deceit that allowed buyers to successfully detect deception were not analysed. Furthermore, this selling scenario used face-to-face interactions, making it unclear whether deception may be easier to detect when the interactions occur in person, compared to online. The probability of deception occurring within a negotiation is influenced by the negotiator's emotional state (Gasper & Schweitzer, 2013), as well as the emotional state of their counterpart, which may in fact be contagious (Methasani et al, 2017). Contextual elements that can increase deception are "one shot" non-recurring transactions (Eklinder-Frick & Åge, 2016), or anonymous or impersonalized relations between negotiators (Kersten et al., 2002). While becoming more frequent for obvious reasons of convenience, particularly during the COVID-19 pandemic, dematerialized negotiations, such as those conducted by email, combine all these contextual elements that encourage deceptive exchanges (Valley et al, 1998; Gino & Shea, 2012) .

Deterring Rather Than Detecting Deception in Negotiations

Because dematerialized negotiations tend to foster deceptive exchanges, while also limiting the ability to detect potential signs of deception, the present paper proposes an

alternative strategy. Rather than assuming that deception is an integral part of business negotiations and focusing on its detection, we test whether it is possible to deter negotiators from producing deceptive statements in the first place. In other words, we argue that carefully worded questions and statements can make negotiators more honest and transparent during their exchanges. We propose this can be achieved by manipulating a negotiator's cognitive dissonance and their perception of risk.

First proposed by Festinger (1957), cognitive dissonance theory has shaped our understanding of the relationship between cognition, perception, emotions, and motivation (Harmon-Jones et al., 2009). When an individual holds two pieces of information that are related but contradictory, this can create a state of mental discomfort known as cognitive dissonance. When faced with such internal inconsistencies, individuals tend to actively attempt to lower the level of their dissonance (Bendersky & Curhan, 2009). When considering deception, individuals can either lower their cognitive dissonance by finding internal justifications for their deception, e.g., "If I don't get the best price for this then we won't meet our monthly targets", or by actually modifying their behaviors, i.e., reducing the level of deception or telling the truth instead.

In the context of negotiation, creating conditions where an individual feels dissonance when they consider lying to their counterpart could potentially deter them from engaging in deception in the first place (Dissonance factor). This technique is already used in other fields; for example police officers are often advised to build a rapport with suspects and thus increase their cognitive dissonance if they are considering lying (i.e., building rapport can make the suspect implicitly think "because we have a connection, it is a little more conflicting to me to start lying now and breach our trust"), thus obtaining greater cooperation from them (Walsh & Bull, 2012; Vanderhallen & Vervaeke, 2014; Vallano et al., 2015). We postulate one way to accomplish this

in business negotiations contexts is by utilizing a “labeling” (also called "attribution” or “name-calling”) technique (Miller et al.,1975; Strenta & DeJong, 1981; Vallacher & Wegner, 1987, 2012). With the purpose of making people more honest in negotiations, one simple way to implement this technique involves telling a negotiator, at the beginning of the exchange, that you believe them to be “honest and trustworthy”. This label will then be in direct conflict with any deception they are considering, therefore making it potentially more cognitively uncomfortable for them to later lie in the negotiation.

A second way to deter negotiators from lying is through increasing the perceived risk that their counterpart will not believe the lie they are considering (Risk factor). In a business negotiation, where people could be tempted to lie about their floor price in order to justify their selling price, a buyer can attempt to increase this Risk factor in the seller’s mind by mentioning, from the beginning, that they already have an attractive offer from a competitor at a given low price. This well-documented technique, consisting of creating a fixed point of reference early in the negotiation, is known as “anchoring”, and shapes expectations by defining a lasting basis for the negotiation (Chandrashekaran & Grewal, 2006; Beggs & Graddy, 2009; Bokhari & Geltner, 2011; Nasiry & Popescu, 2011; Scott & Lizieri, 2012).

For example, imagine a supplier (supplier A) planning to tell a prospective client that they cannot lower the selling price of a product below $500 because they would then make no profit, when in fact it only costs them $100 in total to manufacture. Would they be less tempted to do so if the potential client started the conversation by announcing that they found the exact same product priced at $400 from supplier B? Such a statement from the prospective buyer, especially if realistic, would make the supplier think that the lie they were considering (making

no profit if selling below $500) has a high risk of simply not being believed. Therefore, the buyer would be more likely to avoid producing such a deceptive statement in this context.

Taking into account what we have called the Dissonance and Risk deterrence factors, we thus hypothesized the following during buyer/supplier negotiations conducted via email:

Hypothesis 1: Labeling negotiators as "honest and trustworthy" at the beginning of the exchange will increase the dissonance induced by engaging in lying, and will therefore decrease the proportion of negotiators providing deceptive statements compared to the no labeling conditions (Dissonance factor).

Hypothesis 2: Providing a low anchor price at the beginning of the exchange will increase negotiators' perceived risk of being disbelieved if they engage in lying, and will therefore decrease the proportion of negotiators providing deceptive statements compared to the high anchor price and no anchor price conditions (Risk factor).

Hypothesis 3: Increasing both the Dissonance and Risk factors felt by negotiators will decrease the proportion of negotiators providing deceptive statements to a larger extent than only increasing one of these two factors.

Method

Participants

In line with the number of participants used in similar studies (van Kleef et al., 2004; Giordano et al., 2007; van der Wijst & Krahmer, 2009; Ott et al., 2013), we recruited 490 participants in total, to reach a minimum of 50 participants in each of the eight experimental conditions included in the final analysis. G-power software was used to confirm the sample size. With an assumed medium effect size ($\omega = 0.30$) of the multigroup goodness of fit tests with

alpha = 0.05 and a statistical test power of 1-β = 0.95, a minimum target sample size of n = 243 participants was determined.

Participants were recruited online using the Turkprime platform (Litman et al., 2017), which draws from a pool of millions of respondents. We believed that this would more closely emulate real-world negotiating conditions than simply asking for volunteers from our networks or students. We set the inclusion criteria so that only participants living in the USA, with an approval rating of over 90% on the platform, could participate in the study, ensuring that they had a verifiable track record for completing surveys. Accepted submissions were rewarded with $1.2 USD. Thirty-two participants were excluded from the analysis for failing to understand the instructions of the study (e.g., by proposing to sell at a loss or increasing their selling prices per unit while they were offered larger volumes of order).

In total, 458 participants were included in the analysis (249 males and 209 females). Subjects belonged to a wide range of ages and professions (see Tables 2.1 and 2.2), with the majority of participants in the 25-34 (43.0%) and 35-49 (33.6%) age groups, and holding positions in management (25.3%), sales (15.4%), or services (14.5%). On average, participants spent 14.8 minutes completing the online survey.

Table 2.1

Participants by Age Group

Age group	Percentage of participants
18 - 24	13.3
25 - 34	43.0
35 - 49	33.6
50 - 64	9.2
> 65	0.9

Table 2.2

Participants by Occupation

Occupation category	Percentage of participants
Management, professional, and related	25.3
Sales and office	15.4
Service	14.5
Unemployed	10.3
Student	4.1
Construction, extraction, and maintenance	3.7
Government	3.2
Production, transportation, and material moving	3.0
Retired	1.4
Farming, fishing, and forestry	0.7
Other	18.4

Procedure

The procedure was informed by discussions with professional buyers from an international firm. Purchasers within the firm usually discussed quotes via email with the six suppliers operating in their market, and tried to determine whether the suppliers were bluffing about their floor price.

The present study was an online experiment designed to replicate the conditions of real-world negotiations as closely as possible. This took the form of a series of emails capturing the conversational interaction when a purchaser sends out an email to suppliers, setting out their needs and asking for a quote. Participants played the role of a supplier of plastic. They were informed that, because plastic was derived from oil, its market price fluctuated daily. At the start of the experiment, participants were provided with some background information about their business and the prevailing market conditions, including:

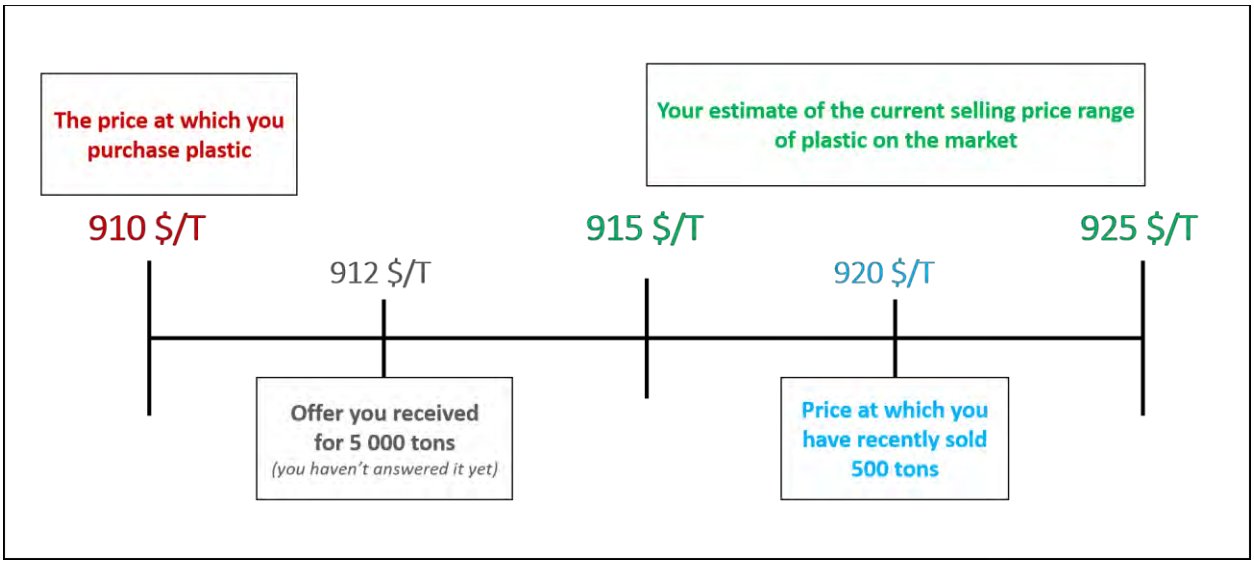
- Typical price range in the current market conditions ($915 to $925 per ton of plastic, depending on the volume sold);
- The fact that there were five other plastic suppliers on the market;

- Details of recent deals, providing reference points. More specifically, they recently sold a small volume of plastic at $920 per ton for 500 tons, and received a proposal they have not answered yet from a buyer at $920 per ton for a very large volume of 5,000 tons;
- Their own purchasing price of plastic, providing a floor price of $910 per ton;
- A table laying out the level of profit made according to price agreed and volume sold, as calculated by the formula: Total Profits = ([Selling Price in $/ton] - 910) * (Volume of plastic sold).

The background brief informed participants that they were playing the role of a supplier receiving a request for a quotation from a buyer of a dairy company that required plastic to manufacture its yogurt pots. Participants were informed that when seeking a deal, companies often resort to multiple suppliers to fulfill its needs, but they did not know whether this buyer had contacted any of the other suppliers competing in the market for this deal, and if so, what prices the other companies quoted. Participants were also informed that the company seeking the deal was of course unaware of the purchase price the supplier paid for plastic (see Figure 2.1). Using this information, participants were asked to try to maximize their profits on the deal.

Figure 2.1

Summary of Market Conditions provided to Participants



The experiment gave participants a series of emails that were worded identically apart from the initial email, which we manipulated to give eight different variants across eight different experimental conditions (Table 2.3). Participants were randomly allocated to an experimental condition. The initial email received by the participants either included a pro-social label stating that the participant had the reputation of being an "honest and trustworthy supplier" or did not, as well as one of three anchor prices ($915 / $910 / $905 per ton). The control group had no label or anchor price. We selected anchor prices that gradually increased the Risk deterrence factor, described in hypothesis 2. The levels selected covered $915 per ton, which was the low range of present market conditions; $910 per ton as the breakeven point; and $905 per ton where the participant, as a supplier, would lose money and could not consider matching this price. Thus there were two levels for the independent variable "Dissonance" (labeling and no labeling) and four levels for the independent variable "Risk" (anchor prices of $915, $910, and $905, as well as a no anchor price control condition), resulting in eight experimental conditions.

Table 2.3

Initial Email received by Participants in each Experimental Condition

#	Condition	Initial email
1	No labeling No anchor price	Hi [Name of participant], My company is looking to purchase plastic now to cover our needs for the next quarter. Can you make me an offer for a purchase of plastic that can reach up to 2500 tons? Best regards, Julien Verdy. Purchase Manager
2	Labeling No anchor price	My company is looking to purchase plastic now to cover our needs for the next quarter. *I am contacting you today because you have the reputation of being an honest and trustworthy supplier, offering the lowest and most competitive prices on the market.* Can you make me an offer for a purchase of plastic that can reach up to 2500 tons?
3	No labeling $915 anchor price	My company is looking to purchase plastic now to cover our needs for the next quarter. Can you make me an offer for a purchase of plastic that can reach up to 2500 tons? *FYI - we have already received a first offer from a supplier at $915/ton.*
4	No labeling $910 anchor price	Same as condition 3, but with a $910 anchor price.
5	No labeling $905 anchor price	Same as condition 3, but with a $905 anchor price.
6	Labeling $915 anchor price	My company is looking to purchase plastic now to cover our needs for the next quarter. *I am contacting you today because you have the reputation of being an honest and trustworthy supplier, offering the lowest and most competitive prices on the market.*

#	Condition	Initial email
		Can you make me an offer for a purchase of plastic that can reach up to 2500 tons? *FYI - we have already received a first offer from a supplier at $915/ton.*
7	Labeling $910 anchor price	Same as condition 6, but with a $910 anchor price.
8	Labeling $905 anchor price	Same as condition 6, but with a $905 anchor price.

The online survey built up a specific exchange: whatever price-per-ton the participant initially gave, the buyer twice sought a better price, firstly arguing that he had received a better offer from another supplier (without specifying the corresponding price), then stating that he could double the volume ordered and take the participant as his sole supplier for the deal if the supplier could make a more competitive price-per-ton offer. Ultimately, the survey delivered the experiment's target final email, which was the only one analysed in the study. Participants answered each email by writing their response email in free format. Their answer to the final target email was categorized according to predetermined criteria, and the percentage of emails within each category formed the dependent variable for this study. This target email was a particularly provocative message from the buyer that asked for sensitive business information, namely the purchase price the supplier paid for the product. It stated:

> *"Thank you for your last offer. To be totally transparent with you, we are not yet at the price level required by my hierarchy, given the current market conditions. In order for me to understand your constraints, could you tell me the price at which you yourself purchase the plastic? I think it will make it easier for us to see if we can hope reaching an interesting deal for both of us, or not."*

Coding and Analysis of the Dependent Variable

In response to the final target email of the study, participants gave answers that fell into one of three categories. They either gave their true purchasing price ($910 per ton), a false purchasing price (any price >$910 per ton), or dodged the question, for example, by saying words to the effect of "this is unfortunately confidential information I cannot share, but rest assured I have offered our best possible price on this deal."

For each experimental condition, we counted the number of participants falling into each of the three possible types of answers and converted these to a percentage of the total number of participants for the experimental condition.

Results

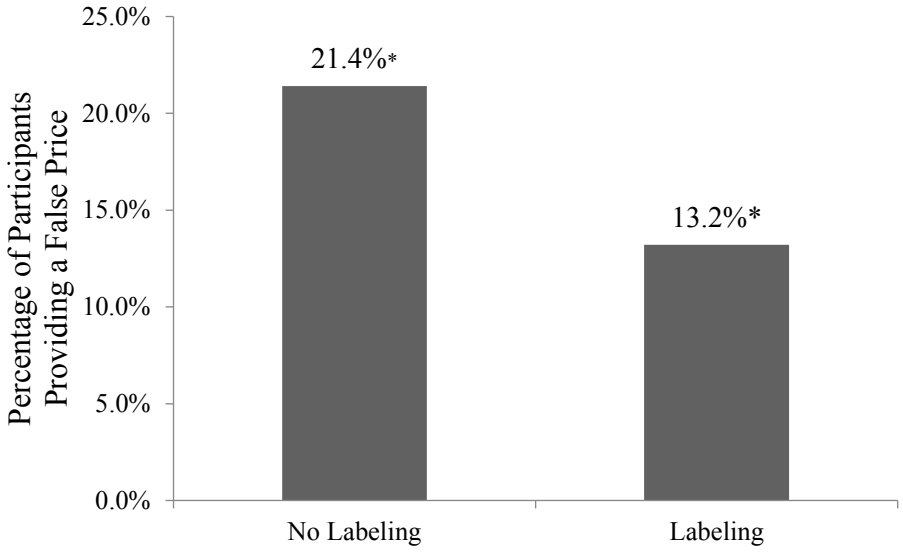
As the dependent variable was categorical, and the data analysed in terms of frequencies for each condition, hypotheses were tested using a test of two proportions to assess the proportion of liars and truth tellers between conditions. For the test of two proportions, we used the chi-square test of homogeneity on SPSS software. Before analysing the effects of the Dissonance and Risk factors on the percentages of various responses, we perform the chi-square test for homogeneity to ascertain whether the experimental groups were correlated to the replies. In this case, we have eight experimental groups and three response types. The chi-square value was 42.9, with a p-value of .003. This implies that there was a correlation between the proportion of the three response types and the experimental groups. In addition, we performed a linear model analysis with interaction effects to explore combined effects. The subsequent sections will explore this correlation in detail.

High Level Effect of the Dissonance factor

To test whether using a labeling sentence as a means to increase the Dissonance factor would make participants more honest (Hypothesis 1) at a high level of analysis, data were subjected to a test of two proportions. A total of 458 participants were randomly assigned to one of eight experimental conditions (Table 2.3), half of which included a labeling sentence. At a high level across the conditions, this corresponded to 243 participants being assigned to a condition including a labeling sentence and 215 participants assigned to conditions with no labeling. Among the 215 participants assigned to a No Labeling condition, 46 participants (21.4%) gave a false price in response to the sensitive target question about their purchasing price, compared to 32 among the 243 (13.2%) in the Labeling conditions, a statistically significant difference in proportions of 8.2%, $p = .02$, $h = .22$ (see Figure 2.2).

Figure 2.2

Percentage of Participants Giving a False Price by Dissonance Factor

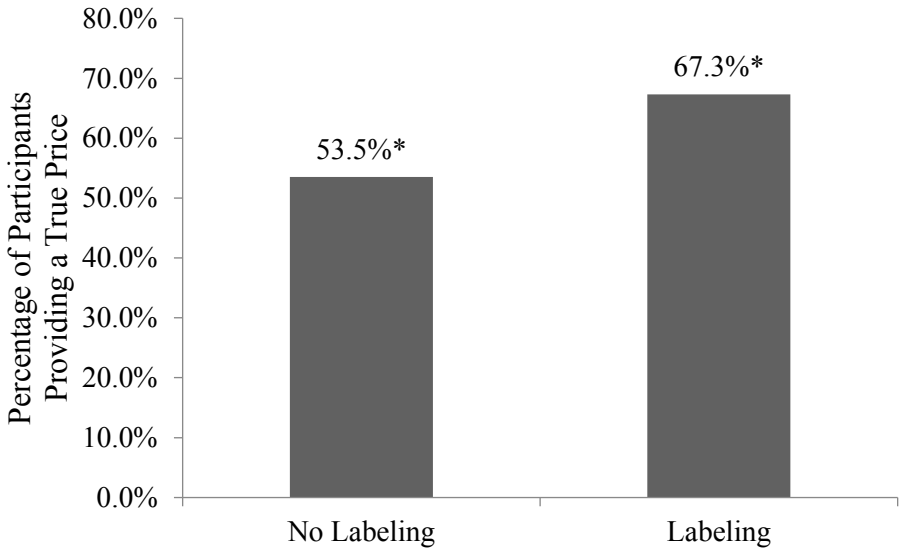


*p = .02

Analysing only the participants who did not dodge answering the target email (i.e., only those who gave a price in response to the target email, whether true or false) led to similar results, with 53.5% of the 99 participants in the No Labeling conditions revealing their true purchasing price, compared to 67.3% of the 98 participants across the Labeling conditions, a statistically significant difference in proportions of 13.8% (p = .048, h = .28; see Figure 2.3).

Figure 2.3

Percentage of Participants Giving a True Price by Dissonance Factor After Excluding Participants Who Dodged the Question



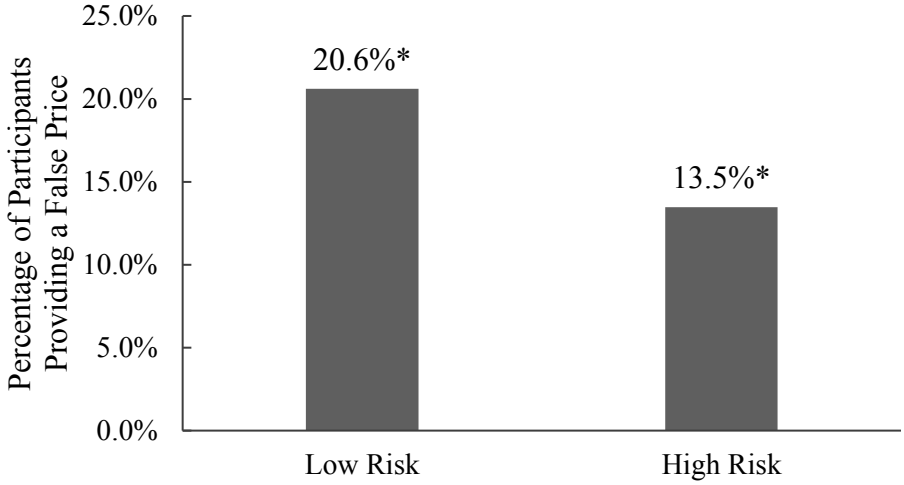
*p = .048

High Level Effect of the Risk factor

To test whether increasing the Risk factor would make participants more honest (Hypothesis 2) at a high level, conditions were grouped in a balanced manner. Conditions with no anchor price or with a reasonable anchor price ($915 per ton) were grouped into a "Low Risk factor" group, while those with anchor prices that were impossible to match ($910 and $905 per ton) were grouped into a "High Risk factor" group. Results showed that 20.6% of the 228 participants in the Low Risk factor group gave a false purchasing price, compared to 13.5% of the 230 participants in the High Risk group (see Figure 2.4), a statistically significant difference in proportions of 7.1% (p = .042, h = .19).

Figure 2.4

Percentage of Participants Giving a False Price According to Risk Factor Level

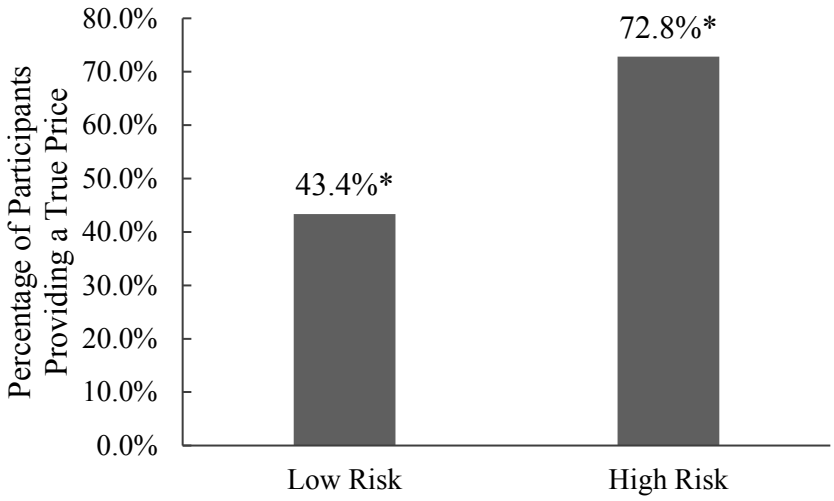


*p = .042

Of the participants who did not dodge answering the target email, showed that 43.4% of the 83 participants in the Low Risk group revealed their true purchasing price, compared to 72.8% of the 114 participants in the High Risk group (see Figure 2.5). This gave a statistically significant difference in proportions of 29.4% ($p < .001$, $h = .61$).

Figure 2.5

Percentage of Participants Giving a True Price According to Risk Factor Level After Excluding Participants Who Dodged the Question

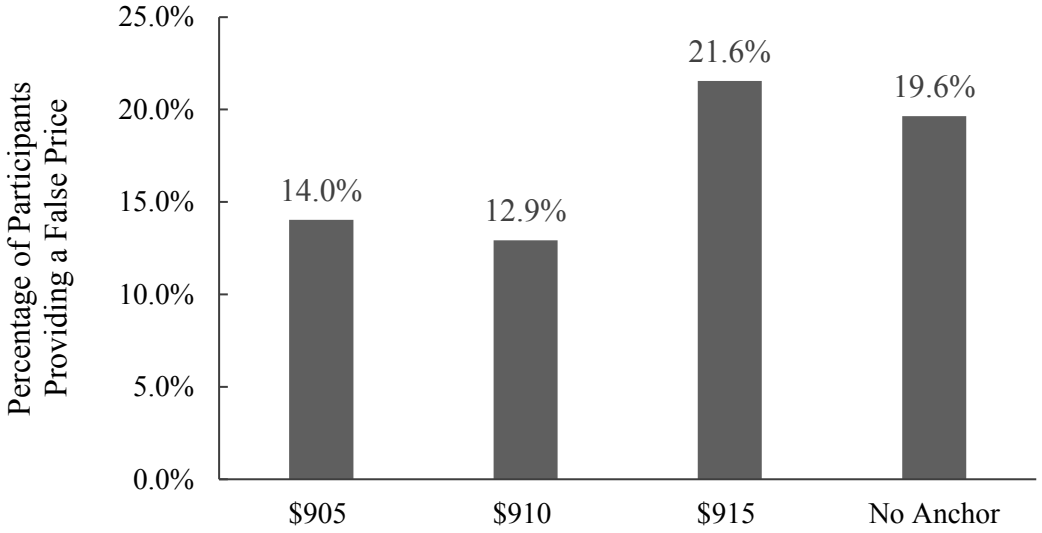


*$p < .001$

We also analysed Risk factor by level of anchoring (see Figure 2.6). Compared to the control group, the anchor price levels of $905, $910, and $915 were all non-significant, with p-values of .26, .17, and .72, respectively. However, when collapsing the data for High/Low Risk, we see a significant difference ($p = .042$; see previous results above). This is a result of the increased sample size in the proportion test.

Figure 2.6

Percentage of Participants Giving a False Price According to Different Anchor Prices



Note. There were no significant differences between conditions.

Combined Effects of the Dissonance and Risk factors

The combined effects of Dissonance and Risk factors can be visualized through interaction plots, which show the false price % vs. dissonance at the different Risk factor levels. The combined effect of the two was quantified using a general linear model with interaction terms (Table 2.4). The interaction term was not statistically significant. There was, however, an additive effect as will be shown later in this section.

To evaluate the combined effects of the Dissonance and Risk factors (Hypothesis 3), data were first analysed at a high level (Labeling/No Labeling x Low/High Risk). As Figure 2.7 shows, a significant difference was found in the percentage of participants giving a false price in response to the target email between the No Labeling/Low Risk group (26.7% of the 101 participants) and both the Labeling/Low Risk group (15.7% of the 127 participants, $p = .042$, $h = .27$) and the Labeling/High Risk group (10.3% of the 116 participants, $p = .002$, $h = .43$).

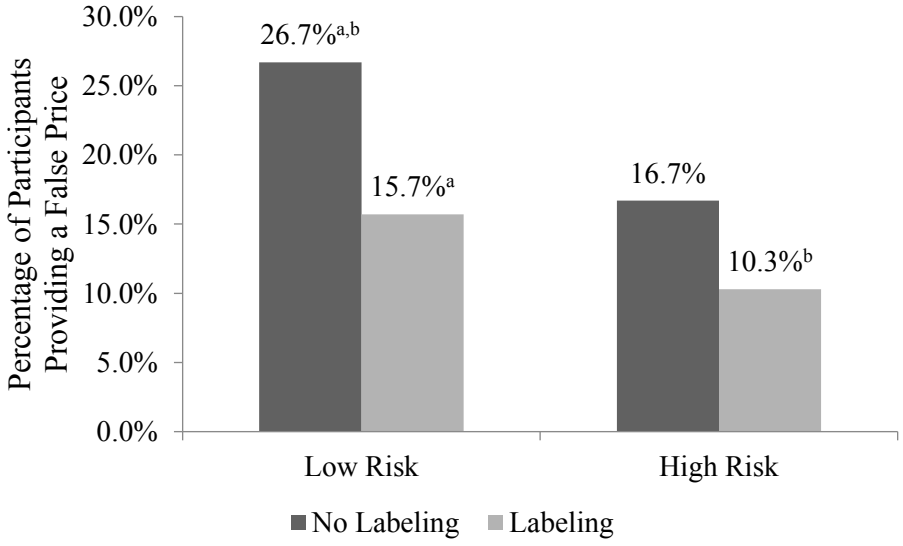
Table 4

General Linear Model Results for Fitting Percentage of Participants Providing a False Price by Labeling/No Labeling and Risk factor as Predictor Variables

$F(3,4) = 8.49$, $p = .033$, adjusted $R^2 = 0.762$			
Coefficient	**Value**	**t value**	**p-value**
Constant	0.175	14.99	<.001
Labeling	-0.044	-3.752	.020
Risk	-0.038	-3.231	.032
Labeling*Risk	0.011	0.973	.386

Figure 2.7

Percentage of Participants Giving a False Price According to Low/High Risk Factor and with/without Dissonance Factor

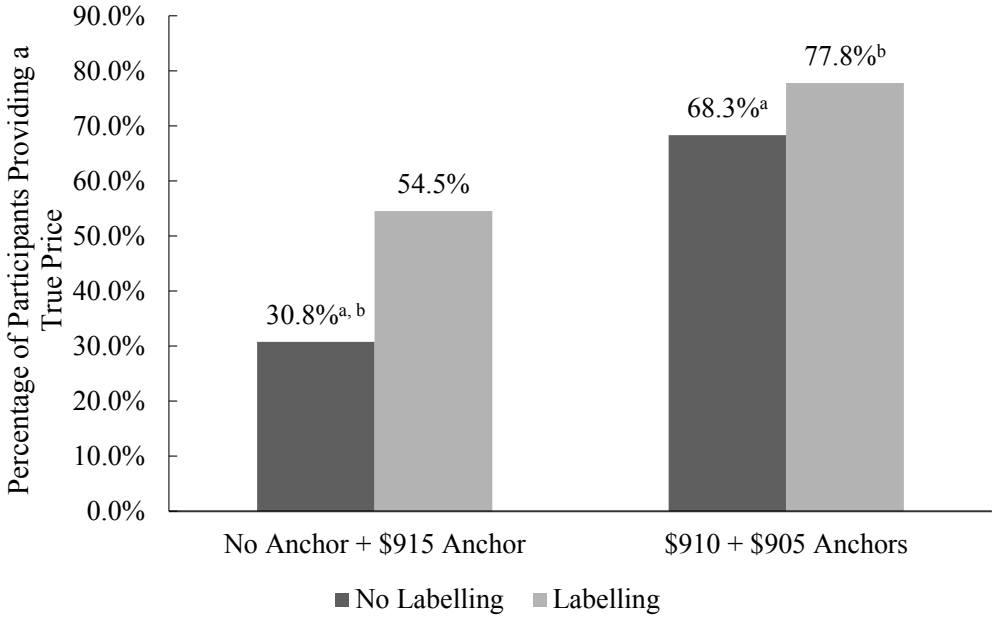


Note. Same index letters indicate a statistically significant difference between conditions ($p < .05$).

Figure 2.8 shows the stronger differences between these groups once participants who dodged the question and did not provide a purchasing price in response to the target email were excluded from the analysis. A significant difference was found in the percentage of participants giving a true price between the No Labeling/Low Risk group (30.8% of the 39 participants) and both the No Labeling/High Risk group (68.3% of the 60 participants, $p < .001$, $h = .77$) and the Labeling/High Risk group (77.8% of the 54 participants, $p < .001$, $h = .98$). The effect sizes for these last analyses in particular ($h = .77$ and $h = .98$) met or exceeded (Cohen, 1988) convention for a large effect ($h = .80$).

Figure 2.8

Percentage of Participants Giving a True Price According to Low/High Risk Factor and with/without Dissonance Factor Excluding Participants Who Dodged the Question



Note. Same index letters indicate a statistically significant difference between conditions ($p < .05$).

At the experimental condition level (Table 2.5 and Figure 2.9), a significant difference was found in the percentage of participants giving a false price (among all participants) between the Labeling/$905 Anchor condition (8.2% of the 61 participants in this condition provided a false price) and both the control No Labeling/No Anchor condition (26.0% of the 50 participants in this condition provided a false price, $p = .011$, $h = .49$) and the No Labeling/$915 Anchor condition (27.5% of the 51 participants in this condition provided a false price, $p = .007$, $h = .52$).

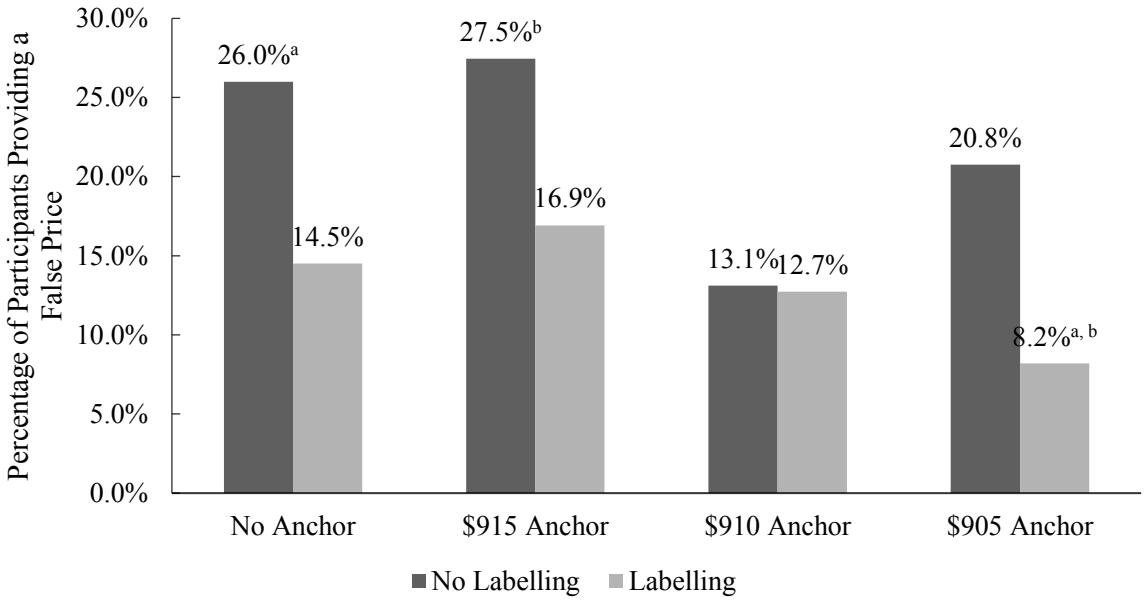
Table 2.5

Distribution of Participant Responses to the Target Email per Experimental Condition

Condition	Number of Participants Dodging (%)	Number of Participants giving True Price (%)	Number of Participants giving False Price (%)	TOTAL
No Labeling/No Anchor	29 (58)	8 (16)	13 (26)	50
No Labeling/$915 Anchor	33 (65)	4 (8)	14 (27)	51
No Labeling $910 Anchor	31 (51)	22 (36)	8 (13)	61
No Labeling/$905 Anchor	23 (43)	19 (36)	11 (21)	53
Labeling/No Anchor	42 (68)	11 (18)	9 (15)	62
Labeling/$915 Anchor	41 (63)	13 (20)	11 (17)	65
Labeling/$910 Anchor	27 (49)	21 (38)	7 (13)	55
Labeling/$905 Anchor	35 (57)	21 (34)	5 (8)	61
TOTAL	261 (57)	119 (26)	78 (17)	458

Figure 2.9

Percentage of Participants Giving a False Price per Experimental Condition



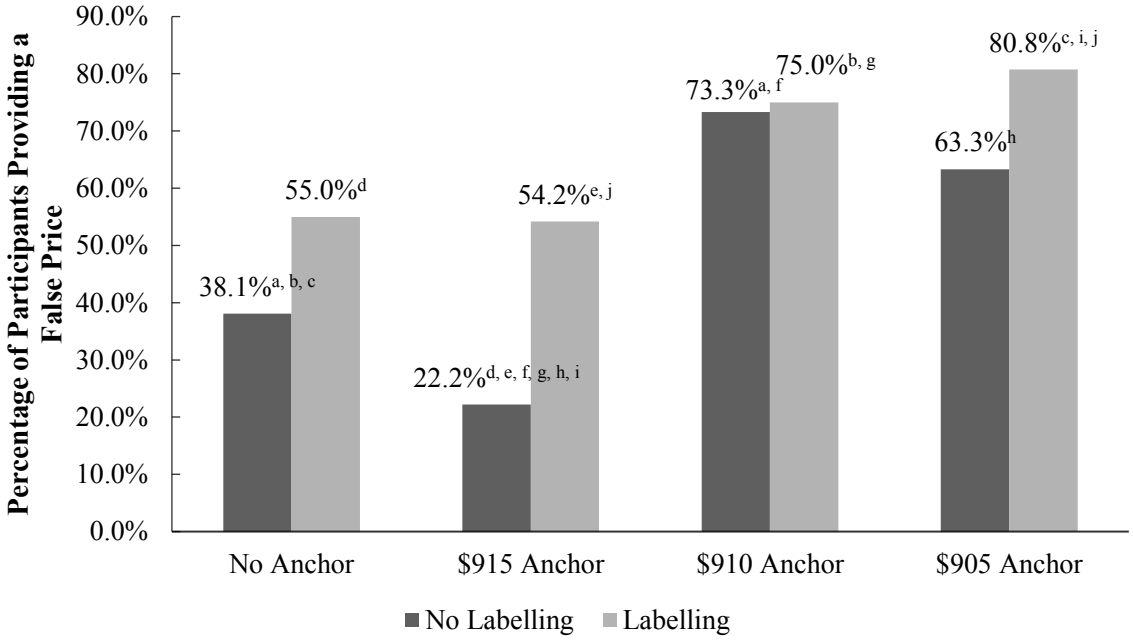
Note. Same index letters indicate a statistically significant difference between conditions ($p < .05$).

Further analysis among only those who gave a purchasing price, either true or false, in response to the target email, led to more numerous significant differences between conditions in the percentage of participants who gave a true price (see Figure 2.10). The control No Labeling/No Anchor condition, in which 38.1% of the 21 participants shared their true purchasing price, was significantly different to the No Labeling/$910 Anchor condition (73.3% of 30 participants, $p = .012$, $h = .73$), the Labeling/$910 Anchor condition (75.0% of 28 participants, $p = .009$, $h = .76$), and the Labeling/$905 Anchor condition (80.8% of 26 participants, $p = .003$, $h = .90$), with large effect sizes. The No Labeling/$915 Anchor condition (22.2% of 18 participants gave the true purchasing price) was significantly different ($p < .05$) to all other conditions except the control No Labeling/No anchor control condition. Finally, the

Labeling/$915 Anchor (54.2% of 24 participants) was significantly different to the Labeling/$905 Anchor condition (80.8% of 26 participants, $p = .044$, $h = .58$).

Figure 2.10

Percentage of Participants Giving a True Price per Experimental Condition Excluding Participants Who Dodged the Question



Note. Same index letters indicate a statistically significant difference between conditions ($p < .05$).

Discussion

In the present experiment, we tested a new approach designed to address the issue of deception in negotiations. Given the difficulty associated with detecting deception in general (Bond & DePaulo, 2006), and in particular for negotiation settings involving non-narrative statements, we proposed an innovative strategy based on trying to *deter* deception, rather than

trying to *detect* it. To operationalize this strategy, we hypothesized that two factors could be used as leverage in the early stages of a negotiation exchange to make people "more honest".

Firstly, we postulated that a negotiator could try to increase the mental conflict felt when their counterpart is considering a lie (Dissonance factor), by labeling them as "honest and trustworthy" at the beginning of their interaction. Our results supported this hypothesis, with a small effect size of the Dissonance factor, namely the four conditions that involved labeling against the four that did not.

Secondly, we hypothesized that negotiators were more likely to produce a deceptive statement if they thought there was a good chance that their counterpart would believe them, and vice versa. Therefore, we postulated that negotiators could make their counterparts more honest by increasing the latter's perception of the risk that they would not be believed if they engaged in deception (Risk factor). Reinterpreting the concept of anchoring (Chandrashekaran & Grewal, 2006; Beggs & Graddy, 2009; Bokhari & Geltner, 2011; Nasiry & Popescu, 2011; Scott & Lizieri, 2012) within our theoretical framework, we assumed that mentioning very competitive alternative offers at the beginning a negotiation was one way of increasing the Risk factor perceived by the counterpart. Our results also supported this hypothesis, with small to medium effect sizes found at a high level of analysis across conditions.

A natural consequence of our first two hypotheses was that combining these two deterrence factors would produce the largest effects on the production of deceptive statements. Comparing the percentage of participants providing a false or true purchasing price in each experimental condition showed trends in the expected directions (see Figures 2.9 and 2.10), especially when comparing the No Labeling/No anchor control condition with the most extreme Dissonance plus High Risk Labeling/$905 anchor condition. Between those two conditions, the

percentage of participants providing a false price decreased from 26.0% (control condition) to only 8.2%, a 3.2 division between the two proportions. Similar results were found when excluding participants who dodged the question and did not provide any true or false purchasing price, with only 38.1% sharing their true purchasing price in the control condition, as opposed to a high 80.8% in the Labeling/$905 anchor condition. Overall, our results showed that increasing the Dissonance factor from "No Labeling" to "Labeling" and/or the Risk factor, from "No Anchor", to a $915, $910, or $905 anchor price, tended to deter lying progressively and decrease the proportion of participants producing deceptive statements. This trend, however, was not linear (see Figures 2.9 and 2.10), and the use of the $910 anchor price, in particular, appeared to override the concomitant use of labeling. We speculate that this result was linked to the fact that the $910 anchor price corresponded to the participants' breakeven point in the current experiment and could thus have boosted the psychological effect of this anchor price.

Practical Implications

In many ways, this research was a simplified version of more complex real-life business negotiations. Indeed, due to the current global uncertainties, dematerialized negotiations have increased, as many exchanges that would previously have been conducted in person are now taking place online. These findings have clear practical benefits for anyone who uses or instructs in negotiations or works with negotiators. Increasing dissonance to lying using an honesty-labeling sentence appears to be easily implemented by real-life purchasers and negotiators in an undetected manner, even in dematerialized email negotiations, as in our experiment. However, the practicability of the use of anchor prices as a means to deter deception is more debatable. In the current experiment, participants played the role of a supplier competing on a market where companies typically sold plastic for between $915 and $920 per ton, but prices could go as low

as $912 per ton for huge volumes of plastic sold. In any case, the supplier purchased the plastic at $910 per ton. Unsurprisingly, our results showed that the effects of anchoring were the most important when offers from competitors of $910 or even $905 per ton, which, in other words, represented offers impossible for the participants to match, were mentioned by the client in their first email. In real life, there may be no problem with mentioning such alternative offers if they exist, but what about the cases where the negotiator has no alternative offer, or the prices are not so unbeatable that they drag the prices to a minimum? Should they resort to "bluffing" and pretend that they have such offers anyway, running the risk that their counterpart will not believe them, or even running the risk of losing their best potential supplier while trying to haggle over a few more dollars?

Limitations and Future Research

The experimental paradigm made the price of plastic the only dimension to be negotiated, as can be the case mostly in "one shot" negotiations. Real-life negotiations, however, are rarely mono-dimensional, and many additional details can be toughly negotiated (e.g., deadlines, guarantees, "goodies", etc.). Furthermore, the emotions of both parties can affect the likelihood of deception (Gasper & Schweitzer, 2013; Methasani et al., 2017), yet the emotions felt during a low-stakes controlled experiment are likely to be subdued compared to the emotional investment of real-life negotiations (Bazerman et al., 2000). Accordingly, follow-up research could explore these additional factors, and researchers should ultimately consider field experiments to assess how well our approach translates into real-life effects.

Additionally, as noted by Zhang et al. (2015), culture may play a role in negotiation, especially regarding establishing trust; thus this could be another variable that affects the results. Given the international nature of many negotiations and modern commerce, it is important to

ensure that the results of this study can be extrapolated to other cultures. Trying to generalize results we found for US-based participants could actually damage negotiations for different cultures with dissimilar approaches to business and falsification during negotiations. Some cultures may see bluffing as an inherent part of negotiations, while others may shun such methods, so businesses may need tailored labels and different anchor price levels.

Conclusion

In summary, the results suggest that our novel approach aiming to deter deception, before trying to detect it, is promising and easy to implement. The simple addition of sentences specifically phrased to increase the cognitive dissonance to lying felt by a negotiator's counterpart, or their perceived risk of not being believed if lying, particularly if combined, can dramatically increase the likelihood of people being more honest. Because the number of negotiations taking place via email has increased, new methods that can be operationalized in dematerialized or more anonymous exchanges are becoming increasingly important. As such, this research expands the literature and recommended guidance for face-to-face negotiations and lie detection strategies.

Chapter 3: Making Job Candidates More Honest with a Theory-Based Form: A Field Study

Most employers report finding lies in resumes (Hayes, 2017), yet employers who receive large numbers of resumes cannot realistically perform systematic background checks. The purpose of this study was to develop a cost-effective way for employers to both reduce and more easily detect deception from job candidates at the pre-interview stage. We collaborated with a recruiting and a background check firm to test a mechanism that addresses resume fraud in recruitment applications. Using a customized application form specifically designed to both *deter* and *enhance detection* of deception in applications, 27 real-life candidates all applying for sales positions were asked to complete the customized form at the pre-interview stage, after submitting their initial free-form resume. A total of 269 elements common to both application forms were verified. Our deception deterrence strategy worked, as inaccuracies were reduced from 23% on the free-form resumes to 11% on the customized application forms ($p < .001$, $d = .9$). The customized form also acted as a way to effectively detect deception, as for the past professional experiences for which the candidates did not provide, upon request, contact information to facilitate our verifications, only 39% contained no false information at all, compared to 77% that were fully accurate when contact information was provided ($p = .007$, $h = .78$). This novel theory-based method to make candidates more honest and facilitate lie detection thus appears to be an efficient and cost-effective way to address resume fraud.

Introduction

In May 2012, after just four months on the job, Yahoo! CEO Scott Thompson agreed to resign after the company's board obtained evidence that he had falsely claimed a computer science degree on his resume (Pepitone, 2012). This is not an isolated case. In fact, in a recent study conducted among 2,575 US recruiters, 75% of employers reported finding a lie on a resume (Hayes, 2017). Among a sample of 300 recent jobseekers, 72% embellished resume details, 61% omitted negative information from a resume, and 31% actually fabricated resume details (Henle et al., 2019). In France, according to the 8th Study on Deceptive Resumes (Florian Mantione Institut, 2017), a large-scale survey investigating both recruiters' and candidates' experiences of deception in job applications, a large percentage of the 339 recruiters who responded suspected that candidates often or always embellish their daily responsibilities (65%), their proficiency in foreign languages (64%), the duration of their professional experiences (61%), their job positions (53%), or their wage (52%). Interestingly, among the 100 job candidates who responded, 85% considered resume embellishment a normal practice.

Misrepresentation of academic and work credentials in application materials is negatively correlated with job performance, so hiring only those who are less deceptive on application materials can increase productivity at the company level (Levashina & Campion, 2009). However, while some methods can assess applicants' general honesty (e.g., social desirability scales [Crowne & Marlowe, 1960] and impression management scales [Paulhus, 1998]), or reduce their tendency to answer dishonestly during routine biodata measures (e.g., the response elaboration technique [RET; Schmitt & Kunce, 2002]), these tests are very time consuming for both the employers and the candidates. Furthermore, while they may give a good indication of a candidate's general tendency to lie, they fail to specifically detect or reduce deception within the

candidates' resumes. The current study addresses this by developing a cost-effective way for employers to both reduce and more easily detect deception from job candidates at the pre-interview stage.

General Theory of Deception

Deception refers to "a successful or unsuccessful deliberate attempt, without forewarning, to create in another a belief which the communicator considers to be untrue" (Vrij, 2008, p. 15). Chapter 1 presented a theory of lie production called the General Theory of Deception (GTD). This framework states that five factors, perceived Benefits, perceived Punishment, perceived Risk, perceived Execution cost, and perceived Dissonance, are weighted against each other whenever an individual is evaluating whether to lie or tell the truth. Research indicates that the benefits of lying can be categorized according to the beneficiary of the lie, i.e., (a) benefits for the self, (b) benefits for others, (c) benefits of preserving a relationship (Hample, 1980; Metts, 1989). Within the GTD, the punishment factor refers to the perceived importance of the punishments/negative consequences in the case of an undesirable outcome of the options of lying, telling the truth, or not answering a solicitation for the truth.

Risk, in this case, refers to the perceived risk of raising suspicions or being disbelieved when lying at the time of the lie, the perceived risk of raising suspicions or being disbelieved despite telling the truth, and the perceived risk of raising suspicions when not answering a solicitation. This subjective evaluation requires Theory of Mind, i.e., the ability to imagine the thoughts, feelings, knowledge, or intentions of others, but also to evaluate how likely one's lies are to be believed by another person (DePaulo, 1992; Gombos, 2006; Talwar et al., 2007).

When considering execution costs, individuals can determine the perceived cognitive (e.g., the fabrication of the lie itself), emotional (e.g., feelings of guilt), or control/inhibition

(e.g., inhibiting body impulses related to lying in order to appear honest) costs. Finally, perceived dissonance refers to the mental conflict faced by people when their words or behaviors are inconsistent with their own past behaviors, actions, statements, ideas, or values.

According to the GTD framework, an individual weighs each of these five factors whenever they are considering whether to lie or tell the truth, according to their subjective perception and assessment of the present and the future, at the time of the assessment. As such, people may make different decisions regarding whether to attempt deception, even in similar contexts. Although the weighing of factors is subjective in nature, it does allow for predictions based on the manipulation of these factors. For example, the likelihood that an individual will engage in deception can be manipulated and reduced by increasing the perceived benefits of telling the truth, decreasing the perceived benefits of lying, increasing the perceived risk of lying, increasing the perceived punishment of lying, etc.

Deception Production in Recruitment

Within the context of job applications, deception could be clear falsifications, such as fabricating previous job experience or academic degrees; or embellishments, which are based on facts, but crucial details have been omitted or falsely elaborated upon. Embellishments can concern dates of employment, such as stating “2015–2016” (i.e., up to two full years) for an actual duration than can be as short as “Dec. 2015 – Jan. 2016” (two full months maximum), degrees (e.g., a candidate implying they obtained a degree from a school they attended but did not actually graduate from), or modifying a job title (from “manager” to “senior partner” for example). Other common deceptions include mentioning businesses that cannot be contacted for background checks or providing deceptive reasons for why they are looking for new employment (“I was not thriving in that position”) to obscure the fact that they were terminated. Such

deception on application materials is widespread and may be on the rise, becoming more pronounced during weak economies (Patel, 2009). Job candidates may not be ethically troubled by engaging in such deceptive behaviors because they feel embellishing a resume is common, necessary, or even expected (Bourdage et al., 2018).

Deception Detection in Recruitment

Employers have struggled to find a cost-effective way to vet employees through the application process. Today, some outsource the process, contracting with recruiting or background check firms to vet potential employees in a lawful manner; others internalize this task, while most do not perform background checks at all. Understandably, employment recruiters who receive large numbers of resumes cannot realistically perform systematic background checks and can therefore be particularly exposed to resume fraud. Paradoxically, while 1/3 of recruiters in the 8th Study on Deceptive Resumes declared they had eliminated a candidate after a background check, 2/3 of recruiters reported not performing any background check at all. Recruiters thus appear to rely, to a large extent, on their ability to detect deception and fraudulent resume claims during the application process, which is not likely to be reliable (Bond & DePaulo, 2006; Ekman & O'Sullivan, 1991).

It can be difficult to apply general deception detection techniques to the recruitment process. For instance, well-documented credibility assessment tools initially developed for forensic applications, such as Criteria Based Content Analysis [CBCA] (Köhnken & Steller, 1988; Steller & Köhnken, 1989) or Reality Monitoring [RM] (Johnson & Raye, 1981), require people to make open-ended narrative statements about events (e.g., in response to questions like "tell me everything you did last Sunday"). Yet in the context of recruitment, job recruiters are typically not interested in collecting such open narratives but instead seek answers to more close-

ended questions (e.g., "what was your salary at your last job?"). Deception in this context is therefore harder to detect, because traditional cues to deception, such as the quantity of details provided by interviewees (Vrij, 2008), are non-applicable.

Deception Deterrence in Recruitment

Given the financial cost and time commitment associated with verifying applicants' information, the recruitment process would greatly benefit from interventions that dissuaded applicants from submitting deceptive resumes in the first place. This deception deterrence strategy is advantageous because having more reliable information in applications (i.e., resumes with less deception) potentially allows recruiters to better select appropriate job candidates for the interview stage, without the need for any additional resources, time, or cost from recruiters. Additionally, a deterrence (vs. detection) approach translates better to real-life recruiting settings, as recruiters are typically not trained to identify subtle cues to deception (DePaulo et al., 2003).

Within recruitment and many other contexts, it is commonplace to ask individuals to sign a document or application to verify its accuracy. Interestingly, evidence shows that signing a document related to financial reimbursements at the beginning, rather than the end after having already potentially produced deceptive claims, makes people more honest (Shu et al., 2012), probably as a result of a greater mental conflict (what the GTD will refers to as the "Dissonance factor", in reference to Festinger, 1957) to lie right after signing. This method could be beneficial for recruiters and would also be very easy for them to implement.

Many companies use self-report measures as part of the recruitment process to gain insight into traits that will be relevant in the workplace, for instance life experience and reactions to stressful situations. Interestingly, there is some evidence to suggest that deception in these measures, also known as faking, is not necessarily detrimental to work performance, as faking

can positively correlate with supervisors' subjective ratings (Ingold et al., 2015). However, it is possible that the same candidates who were capable of faking during the interview could continue to fake while working, resulting in positive subjective job performance scores from managers, but not necessarily success according to objective measures. Concerns therefore remain over the extent to which faking can affect hiring decisions (Rosse et al., 1998) and implications for the construct validity of personality measures in selection criteria (Schmit & Ryan, 1993), leading to efforts to decrease it. A large body of evidence indicates that deception in these measures, also known as faking, can be greatly reduced if participants are given a warning that their responses will be verified, for example by past colleagues (see Dwight & Donovan, 2003 for a review). We refer to this as the "Risk factor" in line with the GTD, i.e., the risk of being discovered making a false claim. Furthermore, the effect of the warning is stronger when it both highlights that deception can be identified and details the consequences of any deception being discovered (Dwight & Donovan, 2003), such as being automatically excluded for consideration. We refer to this as the "Punishment factor". Thus, if applicants are asked to confirm information contained within their resume on a separate form which contains both a warning that background checks will verify the information and a reminder that by law they should be providing accurate information, they will be less likely to include deceptive information.

Another method of deception deterrence which has shown to be effective within both real and laboratory recruitment scenarios is the RET (Levashina et al., 2012; Schmitt et al., 2003; Schmitt & Kunce, 2002). The RET asks applicants to provide further information in support of their answers during self-report measures. For example, for a biodata question such as, "In the last few months, how often have you read about cultures different from your own? (1 = never, 5

= very often)", a required elaboration would typically be, "If you answered 2, 3, 4, or 5, list the cultures and the materials read". Thus, when applicants are forced to justify their response with specific details, it increases their Risk factor, and they are less likely to intentionally exaggerate their self-assessed score to the biodata question. Also, through being forced to engage with the question more deeply, candidates are also likely to remember more accurately, decreasing the probability of unintentional misinformation or omission.

However, these tests are not appropriate for all employment situations. Firstly, the extra time they require to complete can potentially deter desirable candidates from applying in the first place – a particularly important drawback for job offers with a shortage of adequate profiles on the market. Secondly, employers must spend a significant amount of time assessing the results of tests not directly related to the candidates' resume claims. Furthermore, Schmitt et al. (2003) found the effects of the RET technique do not carry over to other questions, i.e., they do not reduce deception, or faking, throughout an application, but only for the questions to which the technique is applied. This means that while personality measures and the RET technique can give an indication of applicants' general tendency to lie, among other potentially interesting personality features, they are unable to detect or deter deception within the applicants' resumes, which is probably more important to recruiters during their first shortlisting step.

The Current Study

Using the aforementioned evidence, we designed and tested a mechanism for *deception deterrence* and *detection enhancement* during the application process. Rather than only trying to *detect* deception on the sole basis of the free-form resumes initially provided by candidates at the pre-interviewing stage, we attempted to make the information candidates provided more honest in the first place by first adding deception deterrence factors to a customized application form.

Finally, the customized application form explicitly asked the candidates to provide relevant contact information at each of their former professional experiences to facilitate potential background checks from recruiters. We thus hypothesized the following:

Hypothesis 1: Including psychological deception deterrence factors (the Dissonance, Punishment, and Risk factors) in the customized application form will significantly decrease the proportion of deceptive information provided by the candidates, as compared to their initial free-form resumes.

Hypothesis 2: Asking the candidates to provide contact information of previous professional experiences for potential background checks (i.e., asking for verifiable information to increase the Risk factor) will make the cues to deception more salient in the candidates' customized applications forms, as compared to their free-form resumes. More specifically, we hypothesize that deceptive candidates will avoid providing such contact information.

Method

Working with a recruiting firm and a background check company, we conducted a study to determine the effectiveness of a customized application form in reducing deception on real-life employment applications. Our goal was to test the deception-reducing effectiveness of a customized application form administered after applicants had sent in their regular free-form resumes, but before they were selected for interviews. The study was conducted in accordance with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. All candidates were informed in the customized application form they signed that, in accordance with labor law, the recruiting firm could proceed to background checks regarding their past experiences.

Through our conversations with the recruiting firm, it was hypothesized that deceptive applications typically include inaccurate information going from stretched dates, embellished job titles or reasons for leaving a previous job, to insinuating a degree was obtained while the candidate did not actually graduate from the mentioned university (e.g., "Columbia University, 2010-2012") or even mentioning non-existing organizations, allowing candidates to make unverifiable claims.

We therefore designed a customized application form that tried to deter deception at both the general and specific level. For example, at the general level we included a sentence at the beginning of the customized application form that asked candidates to tick and sign a Yes / No box certifying that they will provide entirely accurate information in the document, increasing the Dissonance factor when they later consider including inaccurate information. We also reminded candidates at the end of the form, right before signing the document again, the law article requiring candidates to provide information in good faith (increasing the Punishment factor) and asked candidates to provide contact information of their previous job positions for potential background checks (increasing the Risk factor).

At the specific level, we also designed the customized application form to ask for key specific aspects that we hypothesized were prone to contain deceptive elements in free-form resumes. For example, for each professional experience, candidates were asked to mention the corresponding beginning and end dates in a day/month/year format, and provide the *legal* reason for leaving their previous jobs from a list of options (resignation, termination, etc.), rather than from a motivational standpoint. Such direct close-ended or multiple-choice questions would force candidates who provided ambiguously deceptive information in their free-form resumes to choose between rectifying any ambiguity, and thus providing fully truthful information in the

customized application form, or engaging in a less morally acceptable and more mentally conflicting bald-faced lie. For example, a candidate voluntarily mentioning employment dates at the year level in their free-form resume to convey the impression of a full 2-year experience (e.g., indicating "2015–2016", for a December 2015 – January 2016 actual employment period), would be forced, in the customized application form asking for dates in a day/month/year format, to either rectify this ambiguity and provide the accurate employment dates (i.e., becoming "more honest"), or engage in a real falsification lie in the customized application form (e.g., lying by reporting 01/01/2015 – 31/12/2016).

Candidates

All candidates (n = 27; 70.4% male) included in the study were applying for the same job type (sales position), in France. The recruiting firm believed that resumes for sales positions contained the highest rate of deceptive information, which is consistent with the 8th Study on Deceptive Resumes, in which 73% of recruiters believed sales position applications were deceptive, followed by 51% of manager positions and 31% of technical engineer positions.

Procedure

The candidates were asked to send in their initial free-form resume (i.e., the generic resume all candidates prepare for their various job applications), and then asked, in a second step, to complete a customized application form at the pre-interview stage. It was not possible to counterbalance the order in which candidates completed the free-form resume and customized application form because real-life candidates seeking work already have their resumes up to date.

The customized application form contained general deception deterrence factors, including a sentence at the beginning asking candidates to tick and sign a Yes / No box certifying that they will provide entirely accurate information in the document. We predicted this would

increase the Dissonance factor when they later considered including inaccurate information and thus decrease the amount of false information, similar to Shu et al (2012). The end of the form also contained a reminder of the law article requiring candidates to provide information in good faith (increasing the Punishment factor), and asked candidates to provide contact information of their previous job positions for potential background checks (increasing the Risk factor), both of which have been shown to reduce deception in the self-report measures stage of a job application (Dwight & Donovan, 2003). The customized application form was also designed to deter deception by using close-ended or multiple-choice questions specifically addressing the most frequent types of deceptive claims in job applications. Such questions were designed to account for the fact that deceptive candidates tend to favor omission or insinuation lies, rather than falsification lies (Hayes, 2017), a deceptive strategy which is no longer available to candidates when asked such questions.

The recruiting firm collected the candidates' free-form resumes and then sent and received back the customized application form completed by the candidates; they then decided whether or not to propose a face-to-face interview with candidates. Both the free-form resumes and customized application forms were verified by the background check company, which conducted regular telephone checks to the companies and universities mentioned in the applications, in addition to checking available databases (e.g., academic databases of graduated students).

Information regarding candidates' ongoing job positions was excluded from verification because current employers might not be aware that their employee is seeking another job. For each resume, the background check company selected the experiences that they considered the most relevant to the position the candidates were applying to. This led to between two and four

professional experiences checked per resume, and between zero and two for the education section. For each educational experience, the following elements were checked: (1) the existence of the mentioned institution, (2) the dates mentioned, (3) the exact name of the degree or training mentioned, and (4) whether the degree was actually obtained by the candidate. For the professional experiences, the background check firm verified (1) the existence of the mentioned institution/company, (2) the dates the candidates mentioned as their start/end date in the organization, (3) the exact name of the job position the candidate held, and (4) the reason for leaving the job. Importantly, dates were defined as misleading if they were either a) not correct, b) not mentioned at all, or c) contained only the years and not the months of the educational or employment experience.

Analyses

A paired-samples t-test was used to determine whether there was a statistically significant mean difference between the percentage of inaccuracies in the free-form resumes and the customized application forms. As the sample size is less than 30, we also employed bootstrapping to ascertain the statistical significance of the results. Power (type II error avoidance) was also calculated. In addition, we show that the data were normally distributed, so the paired t-test is valid. To compare the proportions of professional experiences containing no deceptive information when contact information was provided with those where no information was provided, we used the Fisher's exact test for proportions (as the number of experiences to reach this criterion was small). We recognize that the data points are not independent in this case as one participant may contribute to more than one deceptive element – each participant provides free-form and custom form resumes and may account for more than one accurate/inaccurate

element on both forms. Nevertheless, we have carried out the statistical test to quantify the differences.

Results

A total of 404 elements from 27 pairs of free-form resumes and customized application forms (i.e., coming from 27 candidates) were included for verification. On average, 74% (SD = 21%) of experiences included for verification could be verified by the background check firm, per candidate. In total, 317 elements (78%) could be verified. Among those, 269 were related to information common to both the free-form resumes and custom applications forms – the difference corresponding to elements related to the legal reasons of leaving a job, a detail specific to the customized application form. The remaining 135 elements were unverifiable because the organizations involved could not be contacted or no longer existed (e.g., companies that had closed since the time of the experience). Two types of information, educational and employment-related, were checked.

Hypothesis 1: Prevalence of Deception in the Free-form Resumes vs. Customized Application Forms

Results show that our deception deterrence strategy worked and hypothesis 1 was supported, as inaccuracies were reduced by half from 23% on the free-form resumes (61 inaccurate elements over the total of 269 verified elements that were common to both forms) to 12.3% on average per candidate (30/269 [11%] in total) on the customized application forms, a statistically significant decrease of -10.9% with a large effect size (95% CI [-15.711%, -6.077%], $t(26) = -4.649$, $p < .001$ ($8.5*10^{-5}$), $d = .9$). The calculation of the power of the one-sided t-test resulted in 0.998 with a difference of -10.9, showing that there was an adequate number of

samples in the study. The Shapiro-Wilk normality test resulted in a value of 0.61, implying that the data were normally distributed, and that the t-test was valid.

To confirm the t-test results, we also employed bootstrapping of the difference data. Resampling with replacement, with the simulation run 1,000 times, we obtain a p-value of $1.1*10^{-6}$. This is a lower p-value than that obtained in the t-test, implying that the decrease is even more significant.

Put differently, about three-quarters (76%; n = 204 elements) of elements checked were correct in both the free-form resumes and the customized application forms. When deception did occur, however, it was more likely to occur on the resumes compared to the customized application forms, with 13% (35/269) of the inaccurate elements in the free-form resumes becoming accurate in the customized application forms, and only 1.5% of the experiences (4/269) being correct on the resumes but falsified on the customized application forms.

As can be seen in Table 1, the majority of the inaccurate information came from dates mentioned by the candidates, representing 84% of the inaccurate information in the free-form resumes (51 date-related inaccurate elements over 61 inaccurate elements in total in the resumes), and 77% (23/30) in the customized application forms.

There was no deception claiming study at a non-existent university. Almost all candidates (92%; 23/25 verified elements) listed the correct name of their educational program in both forms, but fewer were truthful regarding whether they graduated, with 80% of the verified elements (20/25) correctly stating that the candidates received a degree in both forms. In the free-form resumes, 20% of the verified elements (5/25) deceitfully claimed graduation from an institution they did not obtain a degree from, vs. 12% (3/25) in the customized application forms. The difference in these two proportions could not be verified by the Fisher's exact test (p =

0.702), and the z-test for proportions was not utilized due to the small number of samples. Regarding deceit concerning employment, the pattern of deceitfulness found for education resurfaced, with no candidate claiming employment at a non-existent organization and very few experiences (8.5%; 4/47 verified experiences) claiming a false job title in either the resume or the customized form.

In the employment category, reason for leaving a job was only asked for on the customized application forms. We targeted 72 elements for verification but 24 were unverifiable, leaving 48 verifiable experiences. Of those, candidates were truthful at a rate of 88% (n = 42).

Table 3.1

Summary results of deception checks for information that appears in both the free-form resumes and customized application forms

Type of Check		Number of non-deceptive elements in both resume and custom form	Number of deceptive elements on both resume and custom form	Number of deceptive elements on resume but not on custom form	Number of deceptive elements on custom form but not resume	Total number of elements targeted for verification	Total number of unverifiable elements	Total number of elements verified
Education								
	Existence of institution	29	00	00	00	29	00	29
	Dates attended	00	06	19	00	29	04	25
	Name of program	23	01	01	00	29	04	25
	Degree graduation	20	03	02	00	29	04	25
Work Experience								
	Existence of organization	69	00	00	00	72	03	69
	Dates of employment	20	14	12	03	72	23	49
	Job title	43	02	01	01	72	25	47
Totals		204	26	35	04	332	63	269

Note: Total number of verifiable elements in this table = 269 in 27 resumes and customized application forms. The legal reason for leaving a job (72 elements targeted) is not included, as this element was only asked in the customized application form.

Hypothesis 2: Cues to Deception in the Customized Application Form

On the customized application forms, candidates were asked to provide information for contacts who could be called to verify the information provided by the candidates on their employment applications. For each professional experience, the goal was to check four elements as previously mentioned: the existence of the company, the job title, the dates, and the legal reason for leaving the job.

We calculated (as a % of the verified information), how many were false according to the background check company. For example, assuming all four aspects of the professional experience could indeed be checked by the background check company, if only the legal reason for leaving the job was false, the % of false information in the customized application form for this experience would be 1/4= 25%.

As can be seen in Table 3.2, results indicate that for the experiences for which the candidates did not provide contacts to facilitate our verifications, nine experiences out of a total of 23 that could be checked (39%) contained no false information at all (i.e., all aspects verified for the experience were correct). For the experiences for which the candidate did provide a contact: 20 out of 26 (77%) that were checked were fully accurate on all aspects, a statistically significant difference in proportions of .38, (p = .01), and an odds ratio of 5.2 (Fisher's exact test), thus supporting hypothesis 2.

Table 3.2

Summary results of the accuracy of work experience information and contact information provided by the candidates

Contact Information Provided for Background Check	% of Accuracy of Work Experiences	Number of Work Experiences in Accuracy Category	% of Work Experiences among Verified Information	Average % of Inaccurate Information among Verified
No				
	0%	9	39%	
	25%	9	39%	21%(1)
	50%	5	22%	
	Uncheckable	13		
Yes				
	0%	20	77%	
	25%	4	15%	8%(2)
	50%	2	8%	
	Uncheckable	6		
Impossible due to company liquidation		4		
Totals		72		

Notes: (1) (9*0%+9*25%+5*50%)/(9+9+5) = 21%; (2) (20*0%+4*25%+2*50%)/(20+4+2) = 8%.

In other words, not mentioning, when requested, a contact for verification of a professional experience almost doubles the risk of deception being present in some aspects (dates or job title, etc.) of the experience in question. This absence of contact mentioned is thus an interesting cue for detecting deception here.

Additionally, among the experiences that could be checked, those with no contact provided had falsified on average 21% of the information, compared to 8% for the experiences with a verification contact provided.

Discussion

The cost of a bad recruitment can be considerable. According to a 2017 report by The Recruitment and Employment Confederation (Recruitment and Employment Confederation, 2017, p. 29), the total cost of an employee hired at a salary of £42,000, who left after eight months, was equivalent to three times the salary of that person (£132,015). Such bad recruitments could be due, to some extent, to incorrect information provided by the candidates to employers during the application and hiring application process.

Our real-life field study confirms previous findings on the prevalence of deception in resumes, as among candidates applying for a sales position, 23% of the verifiable information in their resumes was inaccurate. This figure might well be underestimated, as 19% (65 elements out of 332) of the information provided in the candidates' resumes was uncheckable (e.g., due to company liquidation since the candidate's professional experience), thus representing an important reservoir for potential additional undetected deception.

Paradoxically, though, the majority of recruiters do not perform any background check (Florian Mantione Institut, 2017), probably due to a lack of internal resources or budget to outsource such verifications. To address this situation, we tested, in collaboration with a recruiting firm and a background check company, the effectiveness of using a customized application form in making candidates more honest in real-life pre-interview applications.

This customized application form was specifically designed with theory-based psychological deception deterrence factors based on the GTD described in Chapter 1. Those included creating a greater mental conflict (dissonance) in candidates tempted to provide inaccurate information, by asking them close-ended specific questions, or making the candidates sign at the beginning of the form as opposed to the end, thus creating a greater mental conflict over providing inaccurate information right after signing. In addition, they were warned that there was a risk any fraudulent claims would be discovered during routine background checks for each of their professional experiences, and reminded of the punishments related to providing inaccurate information in a recruitment application according to the law. Furthermore, in line with the theory behind the RET technique, the close-ended questions asked for specific information regarding claims in the applicants' resumes. These specific questions not only increased the Dissonance factor, but also increased the Risk factor, as those who may have been deceptive through omission would have to either tell the truth or add more deceptive, and verifiable, information, thus increasing the risk of being caught lying. All of these deterrence and detection factors have separately shown effects within studies on personality self-report measures in recruitment (Dwight & Donovan, 2003), or laboratory studies unrelated to employment (Shu et al., 2012); however, this is the first to use them to supply recruiters with a simple and cost-effective method for reducing and detecting deception within resumes.

These deterrence strategies were efficient, with inaccuracies reduced by half in the customized application forms compared to the candidates' initial free-form resumes (11% vs. 23%). The use of such customized forms at the pre-interview stage can therefore be very useful for recruiters, as it allows them to better select the candidates to interview face-to-face, as they base their decision on more accurate information about the candidates.

Interestingly, the customized application forms, besides deterring candidates from providing inaccurate information, also revealed cues to deception that were absent from the candidates' initial free-form resumes. More specifically, background verifications showed that candidates who did not provide contact information for their previous professional experiences were almost two times more likely to have been deceptive about some aspect of those experiences. This suggests that of all the potential information to confirm, recruiters should use such cues (i.e., the absence of contact information for some professional experiences in the customized application form) as the focus of their verifications, as deceptive claims remained at a relatively high level even with the use of our customized application form (11% of inaccurate information remaining). This would be particularly useful if they are on strict time constraints, or have a limited budget for performing background checks with external professional companies. At the interview stage, recruiters can also question those specific experiences in order to detect potential remaining deception from candidates.

This study, like most, is not without limitations. It was conducted in only one country, on one single type of job position that was believed by recruiters to contain the highest rate of deception. As such, the customized application form may not apply to countries or job sectors that use different hiring practices. Second, although there was a reduction in deception on the customized application form compared to the free-form resumes, we cannot be certain whether this was due primarily to one of the techniques or a combination of them. While theoretically interesting, it was not possible for us to assess these techniques separately using multiple groups because verifying these genuine job applications was very time consuming for the background check company, who offered their services free of charge for this study. As such, it would be interesting to measure the separate effects of dissonance, punishment, and risk in future studies.

It is important to note that although the number of applications (27 candidates) was limited due to the time-consuming nature of background checks, the present study included a total of 404 elements for verification (and a total of 317 that could actually be verified). Finally, the present study revealed that most inaccuracies were related to dates, including candidates providing dates at the year level (e.g., "2015–2016") in their free-form resumes. While such date formats were indeed factually inaccurate, in a proportion of cases candidates may not have done it with the intention of deceitfully stretching the perceived duration of their professional experiences, but because they simply believed this was the appropriate date format. However, the majority of the inaccurate information in the customized application forms (which imposed the day/month/year format) concerned dates, thus showing that candidates mostly intentionally lie about dates. Furthermore, as our customized application form did also reduce inaccuracies related to the other less prevalent types of deception in the candidates' resumes (e.g., whether they had obtained an academic degree), we believe the customized form to be a robust and costless means of discouraging candidates from making any kind of fraudulent claim. Future research should expand the present promising results to other cultures and job positions to confirm the effectiveness of using a theory-based custom application form to better deter and detect deception in real-life resumes.

Chapter 4: 'Liars are less detailed' …So what? Comparing two recall instructions to detect deception within-subject

Most deception research provides between-subjects results (e.g., liars give on average less detailed accounts), which might be of limited value for professionals evaluating credibility on an individual basis. This study examines the optimal instructions of a within-subject multiple recalls strategy to detect deception. A total of 110 participants, divided into a Lie and Truth group, were randomly placed into four interview conditions: two Basic report-everything instructions (1), a Basic recall followed by an Open depth instruction (2), a Basic recall followed by the Verifiability Approach & Information Protocol (3), and two recalls with the Verifiability Approach & Information Protocol (4). All recalls were coded for total details and verifiable details. Group (lie, truth) x Recall (first, second) was only significant in condition 3, with truth tellers providing more verifiable details in the second recall than the first. A simple within-subject decision rule was derived, allowing a 76.9% discrimination rate. Professionals can optimally evaluate credibility using two recalls (Basic recall followed by Verifiability Approach & Information Protocol) and observing the evolution of verifiable details.

Introduction

For the last forty years, researchers in the field of deception have been looking for reliable cues to deceit to help practitioners in various fields (criminal justice, police, intelligence, etc.). Among the multiple verbal and nonverbal cues that have been studied over the years, 'quantity of details' has been shown to be one of the most reliable indicators of truthfulness (Vrij, 2008, pp. 258-259). However, the vast majority of studies in the field of deception are between-subject designs, comparing one group of participants making a true statement with a second group making a false statement, typically about past activities. Therefore, the behavioral cues that were subsequently found to be significantly different between liars and truth tellers (e.g., 'quantity of details') are only different *on average* between the two groups.

Consequently, transposing between-subjects results (*'liars are less detailed than truth tellers on average'*) to an individual real-life case (*'is this particular suspect lying to me now or not?'*) might not be possible for professionals making case-by-case assessments. Consider, for example, the following insurance claim. Should it be determined detailed enough to be considered an honest claim?

> *I, the undersigned Miss Juliette Dodue, want to declare the theft of my phone. Indeed, on December 12, while I was doing my Christmas shopping with my cousin, someone took advantage of the rush to tear my cell phone out of my hands and run away with it. I would thus like to be reimbursed or receive another one. Thank you in advance for your efficiency.*

Because such veracity judgments are hard to make in real-life situations based on between-subjects general results, practitioners often ask researchers to develop 'within-subject', rather than between-subject, methods to detect lies. That is, they wish to make a

decision about the veracity of an interviewee's statement while taking into account the substantial individual differences that can exist between people facing similar situations.

To date, few attempts to provide within-subject verbal lie detection tools have been proposed. Historically, Statement Validity Assessment (Köhnken & Steller, 1988; Steller & Köhnken, 1989), or SVA, was the first tool developed by researchers that acknowledged and tried to account for individual differences in speech. SVA consists of four stages (Vrij, 2008, p. 204): (i) a case-file analysis to gain insight into the case; (ii) a semi-structured interview to obtain a statement from the interviewee; (iii) a Criteria-Based Content Analysis (CBCA) that systematically assesses the quality of a statement through the scoring of a set of 19 verbal criteria (e.g., Criteria 3 – Quantity of details), leading to a total CBCA score; and (iv) an evaluation of the CBCA outcome via a set of questions (Validity Checklist). SVA thus attempts to weight the total CBCA score of a statement, calculated in the third step of the procedure, by considering person-specific and/or contextual information that could have influenced the CBCA score (step 4), such as the age or the mental abilities of the person being assessed. However, the few studies that aimed to evaluate the consistency of SVA assessments (Gumpert & Lindblad, 2000; Lamers-Wilkelman, 1999; Parker & Brown, 2000) revealed differences between experts, despite SVA's attempt to provide a systematic non-subjective tool.

More recently, other verbal criteria and tools have been proposed to better discriminate between deceptive and truthful statements, especially if combined with specific instructions to the interviewee. Among those, the Verifiability Approach (VA) is based on the premise that liars (and not truth tellers) face the dilemma of wanting to provide statements with many details to appear truthful, but not details that could be checked by the interviewer, such as through talking to witnesses or reviewing documented activities, and hence contradict their claims (Nahari et al., 2014a). Therefore, the VA posits that truthful statements should

contain more verifiable details than deceptive ones, so specifically asking interviewees to provide verifiable details and informing them that their credibility will be assessed based on the number of verifiable details they provide (an instruction called the 'Information Protocol', or IP) should maximize the difference in verifiable details between truthful and deceptive statements (Harvey et al., 2017; Nahari et al., 2014b). Although relatively under-investigated, and only by a few teams, the VA & IP approaches have shown promising results (Harvey et al., 2017; Jupe et al., 2017; Nahari & Vrij, 2014; Vrij et al., 2016). Furthermore, recent meta-analyses have confirmed that truthful statements contain more verifiable information than deceptive statements, but deceptive statements do not contain more unverifiable elements than truthful statements (Palena et al., 2021; Verschuere et al., 2020). Interestingly, only Palena et al. (2021) confirmed the validity of IP. Both meta-analyses also found significant moderating effects of scenario; that is, the differences in the number of verifiable details between liars and truth tellers were greater when statements concerned a specific event (as opposed to something more general like physical health or occupation), which is logical as events provide the potential for more verifiable details. To account for individual differences, new approaches like the VA have suggested coding verifiable details as a ratio rather than an absolute value, e.g., by calculating the total number of verifiable details divided by the total number of unverifiable details, rather than simply counting the number of verifiable details within a statement. However, while the meta-analyses mentioned above both confirmed this approach can distinguish between liars and truth tellers, it is not ideal for practitioners because the use of ratios makes the tool more difficult to apply, especially in real time. Furthermore, it represents more of a 'normalization' of variables rather than truly making the approach a within-subject method.

In order to develop a within-subject method that could be more easily used in the field by practitioners, we have designed an interview protocol inspired by the Cognitive Interview

(Fisher & Geiselman, 1992), an investigation tool designed to help eyewitnesses retrieve more information from their memory of an experienced event. The Cognitive Interview comprises different mnemonic instructions aimed at enhancing memory recollection compared to a simple free recall of the event. For example, recalling the event in reverse order or from another person's perspective after an initial free recall of the event (Geiselman et al., 1984), or detailing the initial testimony with 'peripheral' details, i.e., information which has not yet been recalled (person, action, object, location, speech). Another aspect, known as Open depth instruction, asks interviewees to report everything, and make regular "image freezes" to describe all the details surrounding the actions (Brunel et al., 2013). The Cognitive Interview has been shown to be more effective in obtaining full and accurate testimony than standard police or structured interviews (Clifford & George, 1996; Geiselman et al., 1986; Ginet & Py, 2001). Besides eyewitness interviews, the Cognitive Interview has also shown promising results in the detection of deception during suspect interviews (Geiselman, 2012), with truth tellers reporting more additional information after the initial free recall than liars (Colwell et al., 2013, Colwell et al., 2007). Opposing results have, however, also been reported (Dukala et al., 2019).

We have adapted the concept of multiple recalls during suspect interviews as a within-subject method and designed an experiment to compare the efficiency of different instructions to discriminate truth tellers from liars in a mock crime setting. We hypothesized that:

Hypothesis 1: in the first recall, truth tellers will provide significantly more details than liars in each condition.

Hypothesis 2: across recall instructions, truth tellers will provide significantly more details than liars, regardless of the recall number.

Hypothesis 3: in the second recall, only truth tellers who receive an Open depth instruction will provide significantly more additional total details compared to the first recall.

Hypothesis 4: in the second recall, only truth tellers who receive VA & IP instruction will provide significantly more additional verifiable details compared to the first recall.

Methods and Materials

Participants

A total of 110 university students (33 males and 77 females) participated in the experiment for course credits. G-power software was used to confirm the sample size. With an assumed medium effect size (f = 0.25) of the multigroup goodness of fit tests with alpha = 0.05 and a statistical test power of 1-β = 0.95, a minimum target sample size of n = 110 participants was determined. The mean age was 20.78 years (SD=2.72).

Procedure

The experiment was conducted in the Psychology department of the University of Toulouse Jean Jaurès, in accordance with the ethical standards of the institution and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The experiment was advertised via student groups on social networks, asking for students to participate in a study about deception detection. Participants arrived at the laboratory in pairs and at a predetermined time. They were then invited to sign a consent form indicating that participation was voluntary and that they could withdraw from the experiment at any time without penalty. Participants then filled a questionnaire about their age, sex, and education level.

The experimenter then gave the following instructions to each pair of participants: ‘For the rest of the experiment, we need a liar and a truth teller. You now have a few minutes

to decide together who will be the liar and who will be the truth teller.' Contrary to laboratory studies in the literature, we chose not to randomly assign the participants to one role as, by definition (Vrij, 2008), deception represents a *deliberate* (not imposed) attempt to mislead another person. We were thus hoping to have more realistic 'profiles' of liars in our experiment, i.e., people more likely to lie in real life and thus exhibit more relevant cues to deceit in our experiment.

After the participants in each pair agreed on their respective roles, the experimenter gave different written instructions individually, based on those used by Nahari et al. (2012). Liars were requested to commit a mock theft of a 10-euro store voucher as part of an experiment about deception. They were told that the voucher was in a purse located in another university building (called 'Olympe de Gouges'), and how to get in the room in question. They were also instructed to make sure nobody witnessed them committing the crime, and that somebody would verify later that the purse was still in place and that the voucher had been taken by them. Truth tellers had to perform their normal business for 20 minutes. They were asked not to stay in one place for more than five minutes; to carry out as many activities as they could, such as buying a coffee, having a conversation with a friend, or visiting a book shop; and not to do anything they may later be uncomfortable discussing. Importantly, truth tellers were asked to perform at least one activity in the building in which the crime was committed.

When the two participants returned to the laboratory after 20 minutes, the experimenter informed them that a theft was committed while they were doing their business: 'Professor [name of the professor], after meeting with students regarding their dissertations in a room meeting in the Olympe de Gouges building, realized that he had forgotten some of his belongings when he left the room. Unfortunately, when he recovered them later, he realized that some things were missing. As you were in the Olympe de Gouges building at the time of

the theft, you are a suspect of the theft. Shortly, you will be interviewed about this. When you enter the investigation room, the interviewer will ask you to explain what you have been doing for the last 20 minutes. The interviewer does not know if you are guilty or innocent of the theft. During your interview, try to be as convincing as possible because the investigator will use your statement, and your statement only, to determine whether you are guilty of the theft that has been committed, or not. Here are some instructions detailing what is expected from you during the interview [the experimenter provides different instructions to liars and truth tellers]. Read them individually for five minutes. Then I will bring you to the waiting room, where you will have ten minutes to prepare for your interview.'

In the written instructions provided to liars, they were first informed of the exact mission truth tellers were given for the last 20 minutes. Then, the instructions said: 'You will soon be asked about this theft. Never admit the theft, under any circumstances. If the interviewer believes that you are innocent, you will receive a 0.5-point course credit and will be able to keep your "booty" [the 10-euro voucher], since you will have succeeded in bluffing the investigator. However, if the investigator believes that you are guilty, you will not receive any course credit. In addition, having been found guilty of theft, you will not only have to give the voucher back, but also have to write a one-page essay explaining what should be, in your opinion, the sanctions applied in real life to students committing theft on campus. The interviewer expects truth tellers to present the activities they have carried out on campus for the past 20 minutes. This interview is your only chance to convince the investigator that you are innocent.'

Truth tellers on the other hand received the following instructions: 'If the interviewer believes that you are innocent, you will receive a 0.5-point course credit. However, if the investigator believes that you are guilty of the theft, you will not receive any course credit. In addition, having been found guilty of theft (although you did not actually commit it), you will

have to write a one-page essay explaining what should be, in your opinion, the sanctions applied in real life to students committing theft on campus. This interview is your only chance to convince the investigator that you are innocent.' Those instructions were designed to mimic real life high stakes situations. In criminal investigations, if a suspect is declared guilty in court, they will end up facing the same sanctions whether or not they are actually guilty of the crime. Likewise, both guilty and innocent suspects in real life have the benefit of being believed by the investigator or judge, but liars typically engage in deception because of a greater potential gain (e.g., a financial gain).

Participants were then randomly assigned to interview room 1 or 2 and brought to the waiting rooms next to their interview room to prepare for the interview. Before the interview, the experimenter verified that the voucher had indeed been taken from the purse by the guilty participant. After ten minutes, two investigators (blind to the participants' statuses) each went out of their interview room, called the participant waiting in front of their room, welcomed them, and started the interview. The interviews were video recorded.

In each interview, the interviewer asked the participant for a first narrative of their activities performed in the last 20 minutes, followed by a second narrative of these activities. The instructions used by the interviewers for the first and second narratives corresponded to one of four sets of instructions corresponding to four experimental conditions, to which participants were randomly assigned.

Instructions to interviewers

For the experiment, interviewers were specifically instructed never to interrupt the participants and to silently wait a few seconds after the participants finished their first narrative to give them the instruction for the second narrative. They were also asked to adopt a supportive nonverbal attitude, through regular smiles and head nods, to encourage the participants to speak more (Bull, 2010; Collins et al., 2002; Fisher, 2010; Mann et al., 2013).

At the end of the two recalls, the investigators wrote their credibility assessment on a slip of paper and then told the participant to open and read it. Participants then had to reveal their true/lie status and give their consent form and questionnaire to the interviewer.

Regardless of the interviewers' evaluations and participants' statuses, the interviewers informed the participants that they would receive the course credit for their participation in the experiment and that they would not have to write an essay. The lying participants who were correctly identified as lying by their interviewer had to return the 10-euro voucher they stole, while those incorrectly identified as truthful were told they could keep it.

Experimental conditions

The statements were collected during interviews conducted by the two investigators. Four experimental conditions were tested, in a two-recall interview setting (see Table 4.1 for the detailed instructions).

In Condition 1, two Basic report-everything recall instructions were used, with the interviewers asking the participants to describe everything they did during the last 20 minutes, in as much detail as possible, twice. In Condition 2, the Basic recall instruction was used for the first narrative, followed by an Open depth instruction inspired by the Cognitive Interview protocol. In Condition 3, the same Basic recall instruction was used for the first narrative, and a combined Verifiability Approach & Information Protocol (VA & IP) instruction was used for the second narrative. In Condition 4, the combined VA & IP instruction was used for both the first and second narrative.

In all conditions, once a participant finished their first narrative, the interviewer asked them to recount their story again, justifying the request with the fact that people usually cannot access and share all the details they remember in just one narrative.

Coding of the statements

Each narrative was transcribed and analysed by two coders, blind to the status of the participant and experimental condition. The coding method was based on that proposed by Nahari et al. (2014a). First, three types of details were scored by frequency: perceptual details (details about what the participants saw, heard, or smelled), spatial details (locations or spatial arrangements), and temporal details (times or sequence of events).

Second, those details were coded according to their verifiability. A detail was coded as verifiable if it was related to a documented or traceable activity, performed with a known person or an identified witness, or if the activity was recorded by a technological means (e.g., a surveillance camera). Last, for each narrative coded, the sum of all the perceptual, spatial, and temporal details were recorded as the 'total number of details' of the narrative, and the sum of all the verifiable details as the 'total number of verifiable details'.

Inter-rater reliabilities between the two coders were measured for both the total number of details and verifiable details via intra-class correlation coefficients (ICC). Out of the 220 total transcripts, 22% were randomly selected for the analysis. The ICCs for the two criteria were excellent: total number of details: ICC = .99, $p < .001$; total number of verifiable details: ICC = .99, $p < .001$.

Table 4.1

Detailed instructions used in each experimental condition

Condition	Description	Recall 1 instruction	Recall 2 instruction
1	Two Basic recalls	Tell me everything that happened during the 20 minutes prior to your arrival in this room. Give me as many details as possible. There are no small details, I am interested in everything!	Thank you for this narrative. It is already very complete. Experience shows that in a first narrative, we cannot tell everything. A second narrative is therefore very useful to access as much information as possible. I am now going to ask you to describe again what happened in the 20 minutes prior to your arrival in this room. Give as much detail as possible. Do not be afraid to repeat yourself, it is normal to repeat yourself because you are going to tell the story again.

Condition	Description	Recall 1 instruction	Recall 2 instruction
2	Basic recall + Open depth	Same as Condition 1, Recall 1	Thank you for this narrative. It is already very complete. Experience shows that in a first narrative, we cannot tell everything. Generally, when telling a story for the first time, people focus on the actions and do not report everything that comes to mind. A second narrative is therefore very useful to access as much information as possible. I am now going to ask you to describe again what happened in the 20 minutes prior to your arrival in this room, as if you were telling your story for the first time. Tell me again all the actions, but focus also on all the details of the scene. To help you, make regular image freezes: describe what happened and regularly stop the scene to describe all the details surrounding the actions. Do not summarize. On the contrary, repeat your story from the beginning and develop it. Do not be afraid to repeat yourself, it is normal to repeat yourself because you are going to tell the story again.

Condition	Description	Recall 1 instruction	Recall 2 instruction
3	Basic recall + Verifiability Approach & Information Protocol	Same as Condition 1, Recall 1	Thank you for this narrative. It is already very complete. Experience shows that in a first narrative, you cannot tell everything. A second narrative is therefore very useful to access as much information as possible. I am now going to ask you to describe again what happened in the 20 minutes prior to your arrival in this room, but this time by including as many verifiable details as possible in your story. By verifiable details, I mean all activities: • documented or traceable (phone calls, cash withdrawal at an ATM, etc.) • made with known person or in the presence of at least one identified witness • recorded by a technological means, like a surveillance camera We know from several studies that those who lie prefer to avoid mentioning details that can be verified, while those who are sincere are not reluctant to provide details that can be verified. I am therefore going to look carefully to what extent the details you provide can be verified to judge the sincerity of your statement. Do not be afraid to repeat yourself, it is normal to repeat yourself because you are going to tell the story again. Please note that we may check some of the information you are going to give.

Condition	Description	Recall 1 instruction	Recall 2 instruction
4	Two Verifiability Approach & Information Protocol recalls	Tell me everything that happened during the 20 minutes prior to your arrival in this room. Give as many verifiable details as possible. By verifiable details, I mean all activities: • documented or traceable (phone calls, cash withdrawal at an ATM, etc.) • made with known person or in the presence of at least one identified witness • recorded by a technological means, like a surveillance camera We know from several studies that those who lie prefer to avoid mentioning details that can be verified, while those who are sincere are not reluctant to provide details that can be verified. I am therefore going to look carefully to what extent the details you provide can be verified to judge the sincerity of your statement. Please note that we may check some of the information you are going to give.	Same as Condition 3, Recall 2

Results

Discriminating participants based on the first recall only

As most of the credibility assessment literature involves trying to discriminate liars from truth tellers based on a single narrative of a past event, we ran Mann-Whitney U tests for each experimental condition to determine whether there were differences in the total number of details/verifiable details between liars and truth tellers, in R1 only. The total number of details was not significantly different between liars and truth tellers in any condition, although it approached significance in Condition 1 (two Basic recalls), $U = 138.00$, $z = 1.84$, $p = .069$ and Condition 2 (Basic recall + Open depth), $U = 159.50$, $z = 1.95$, $p = .050$; but not in Condition 3 (Basic recall + VA & IP), $U = 97.00$, $z = .64$, $p = .55$, or Condition 4 (two VA & IP recalls), $U = 103.00$, $z = 0.95$, $p = .36$.

The differences in verifiable details were also not significant: Condition 1 (two Basic recalls), $U = 93.50$, $z = -.23$, $p = .84$; Condition 2 (Basic recall + Open depth), $U = 135.00$, $z = 1.00$, $p = .37$; Condition 3 (Basic recall + VA & IP), $U = 83.50$, $z = -.059$, $p = .96$; and Condition 4 (two VA & IP recalls), $U = 97.00$, $z = .65$, $p = .55$.

Evolution of the number of details

The evolution of the total number of details between recalls one and two in each condition are displayed in Table 4.2. A three-way mixed ANOVA was conducted to understand the effects of status (true/lie), condition (1 to 4), and recall (R1/R2) on the evolution of the total number of details between the first recall (R1) and second recall (R2). The results of the three-way mixed model are shown in Table 4.3. The significant main effects are status, recall number, and condition, whereas the significant two-way interactions are those between status and recall number, and recall number and condition. These interactions can be visualized through interaction plots (Figures 4.1 and 4.2). The three-way interaction was not statistically significant, $F(3, 102) = .22$, $p = .88$, partial $\eta^2 = .007$. There was,

however, a significant two-way interaction between the recall number and status, $F(1, 102) = 6.63$, $p = .011$, partial $\eta^2 = .061$, corresponding to a medium effect size. Statistical significance of a simple main effect was accepted at a Bonferroni-adjusted alpha level of .025. Pertinent to hypothesis 2, there was a significant simple main effect of status in R1, $F(1, 102) = 5.65$, $p = .019$, partial $\eta^2 = .053$ (small effect size), and in R2, $F(1, 102) = 12.87$, $p = .001$, partial $\eta^2 = .112$ (medium effect size). The mean total number of details was higher for truth tellers than liars in R1, a mean difference of 10.28, 95% CI [1.70, 18.85], $p = .019$; even so more in R2, with a mean difference of 21.45, 95% CI [9.59, 33.30], $p = .001$.

Table 4.2

Comparison of the evolution of the number of details and verifiable details between recalls 1 and 2 in each condition.

Condition	Status	Recall 1		Recall 2		N	F	partial η^2	p
		M	SD	M	SD				
Number of details									
1	Lie	31.43	19.94	31.36	19.72	14	1.65	.060	.21
	Truth	49.43	35.59	56.36	32.77	14			
2	Lie	34.07	16.79	53.33	26.42	15	2.69	.088	.11
	Truth	46.40	16.82	81.27	37.26	15			
3	Lie	35.08	17.29	34.38	18.46	13	2.90	.11	.10
	Truth	44.38	27.89	57.15	38.23	13			
4	Lie	35.46	22.85	38.92	28.28	13	.63	.025	.44
	Truth	36.92	17.04	49.00	41.23	13			
Number of verifiable details									
1	Lie	3.07	5.51	3.79	6.54	14	.41	.016	.53
	Truth	1.79	3.19	1.86	2.45	14			
2	Lie	1.13	1.85	3.60	5.12	15	1.50	.051	.23
	Truth	2.60	4.93	3.27	5.05	15			
3	Lie	2.69	5.54	2.08	3.01	13	6.28	.21	.019
	Truth	1.92*	3.71	6.08*	7.74	13			
4	Lie	6.54	11.52	10.69	15.82	13	1.13	.045	.30
	Truth	6.85	11.79	7.00	7.86	13			

Note: statistics of 2 (Status) x 2 (Recall) mixed ANOVA tests with the true/lie status as between-subjects factor, the recall number as within-subjects factor, and the total number of details/verifiable details as dependent variable are reported for each condition.

* $p = .031$.

Figure 4.1

Interaction Plot of Status and Recall Number with Number of Details as the Dependent Variable.

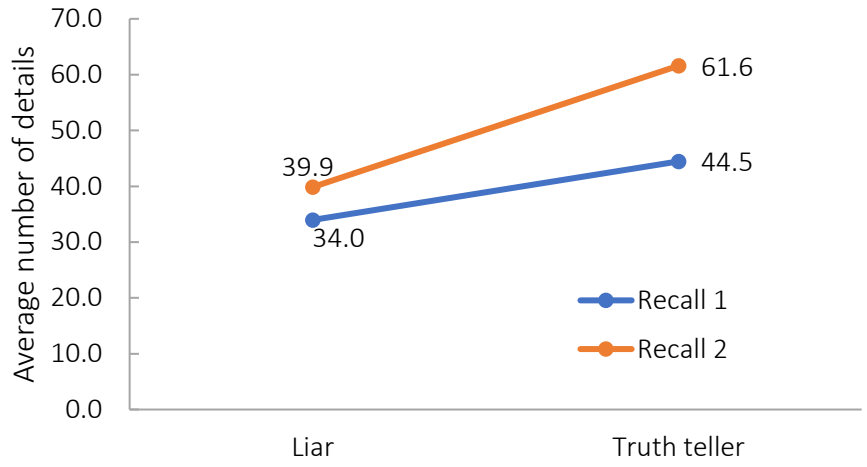
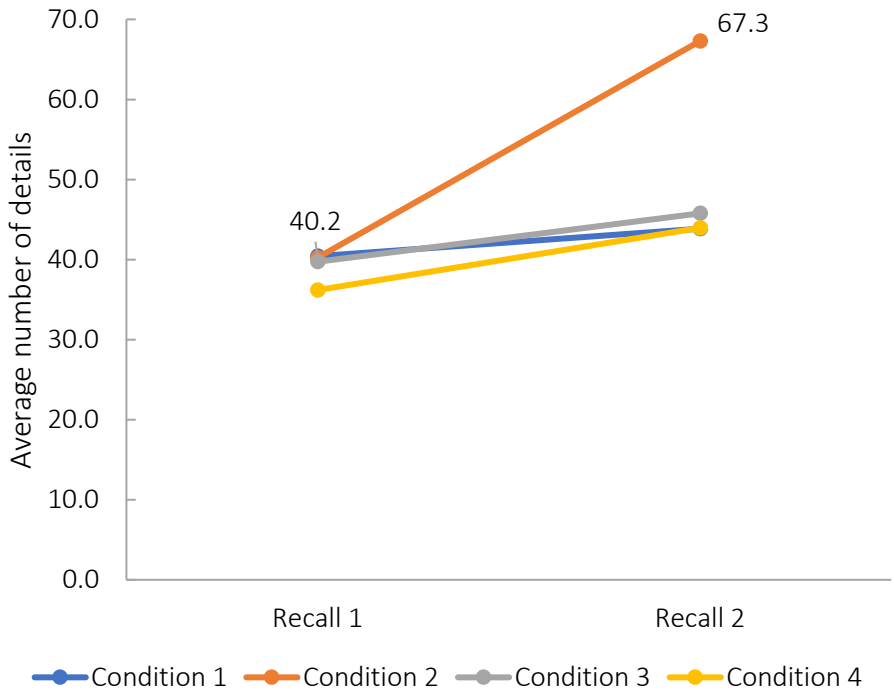


Figure 4.2

Interaction Plot of Conditions and Recall Number with Number of Details as the Dependent Variable.

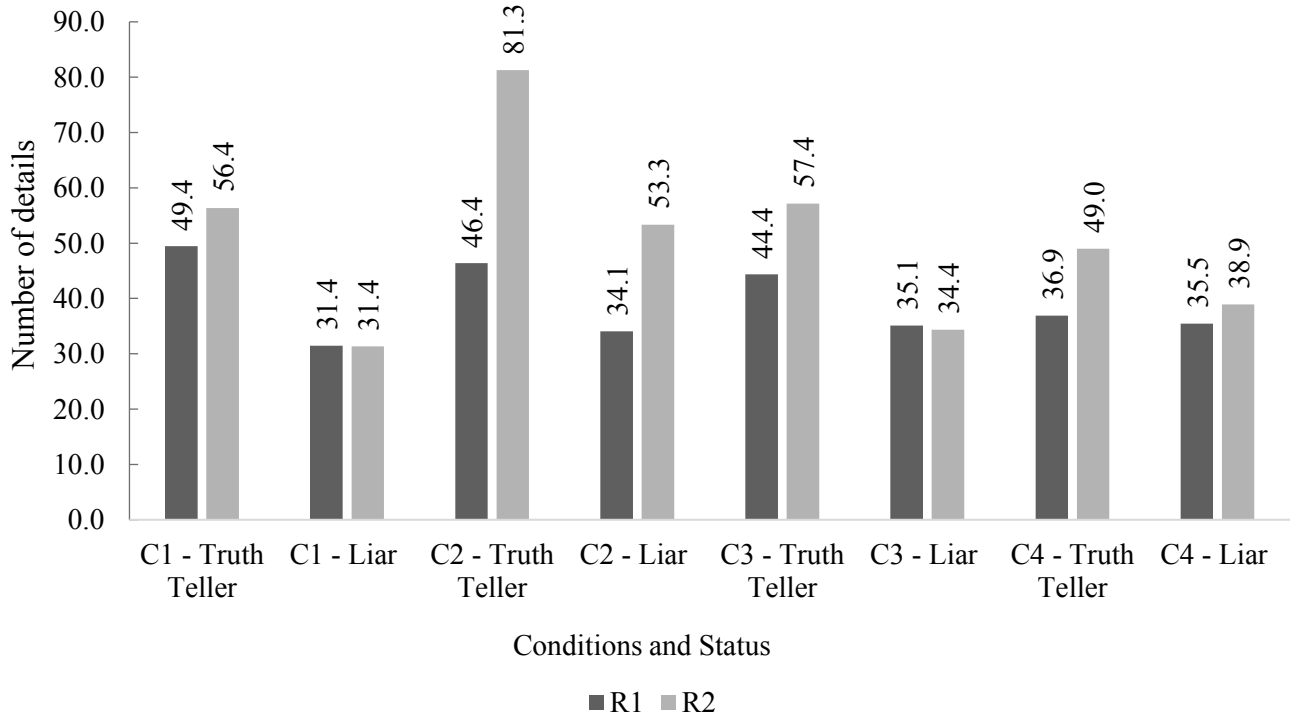


Note: C1 = Two Basic recalls, C2 = Basic recall + Open depth recall, C3 = Basic recall + Verifiability Approach & Information Protocol, C4 = Two Verifiability Approach & Information Protocol recalls.

There was also a significant two-way interaction between recall number and condition, $F(3, 102) = 6.61$, $p < .001$, partial $\eta^2 = .163$, corresponding to a large effect size. A significant simple main effect of condition was found in R2, $F(3, 102) = 3.88$, $p = .011$, partial $\eta^2 = .102$ (medium effect size) but not in R1, $F(3, 102) = 0.21$, $p = .89$, partial $\eta^2 = .006$. In R2, mean total number of details was higher in Condition 2 (Basic recall + Open depth) than in Condition 1 (two Basic recalls), a mean difference of 23.44, 95% CI [1.32, 45.56], $p = .032$; and also higher in Condition 2 (Basic recall + Open depth) than in Condition 4 (two VA & IP recalls), a mean difference of 23.34, 95% CI [.78, 45.90], $p = .038$, as shown in Figure 4.3.

Figure 4.3

Comparison of the Number of Details Recalled by Liars and Truth Tellers in Recalls 1 and 2 in Each Condition



Note: C1 = Two Basic recalls, C2 = Basic recall + Open depth recall, C3 = Basic recall + Verifiability Approach & Information Protocol, C4 = Two Verifiability Approach & Information Protocol recalls.

Evolution of the number of verifiable details

The evolution of the total number of verifiable details between recalls one and two in each condition are displayed in Table 4.2. A three-way mixed ANOVA was conducted to understand the effects of status, condition, and recall on the evolution of the total number of verifiable details between R1 and R2 (see Table 4.4). The significant effects are recall number and condition. Main effect plots in Figures 4 and 5 show the effects of recall number and condition. Condition 4 was found to be significantly higher in the post-hoc Tukey test compared to the other conditions (p-values for comparison with C1, C2, and C3 are .0014, .0011, and .007 respectively).

To refine these results, a follow-up 2 (Status) x 2 (Recall) mixed ANOVA with the true/lie status as between-subjects factor, the recall number as within-subjects factor and the total number of verifiable details as dependent variable was run for each of the four experimental conditions. A significant interaction between status and recall number on the total number of verifiable details was found only in Condition 3 (Basic recall + VA & IP), $F(1,24) = 6.28$, $p = .019$, partial $\eta^2 = .21$, corresponding to a large effect size. Truth tellers in Condition 3 provided significantly more verifiable details in R2 ($M = 6.08$, $SD = 7.74$) than in R1 ($M = 1.92$, $SD = 3.71$, $p = .031$), while the number of verifiable details provided by liars did not differ significantly in R2 ($M = 2.08$, $SD = 3.01$) compared to R1 ($M = 2.69$, $SD = 5.54$, $p = .48$), as shown in Figure 4.4.

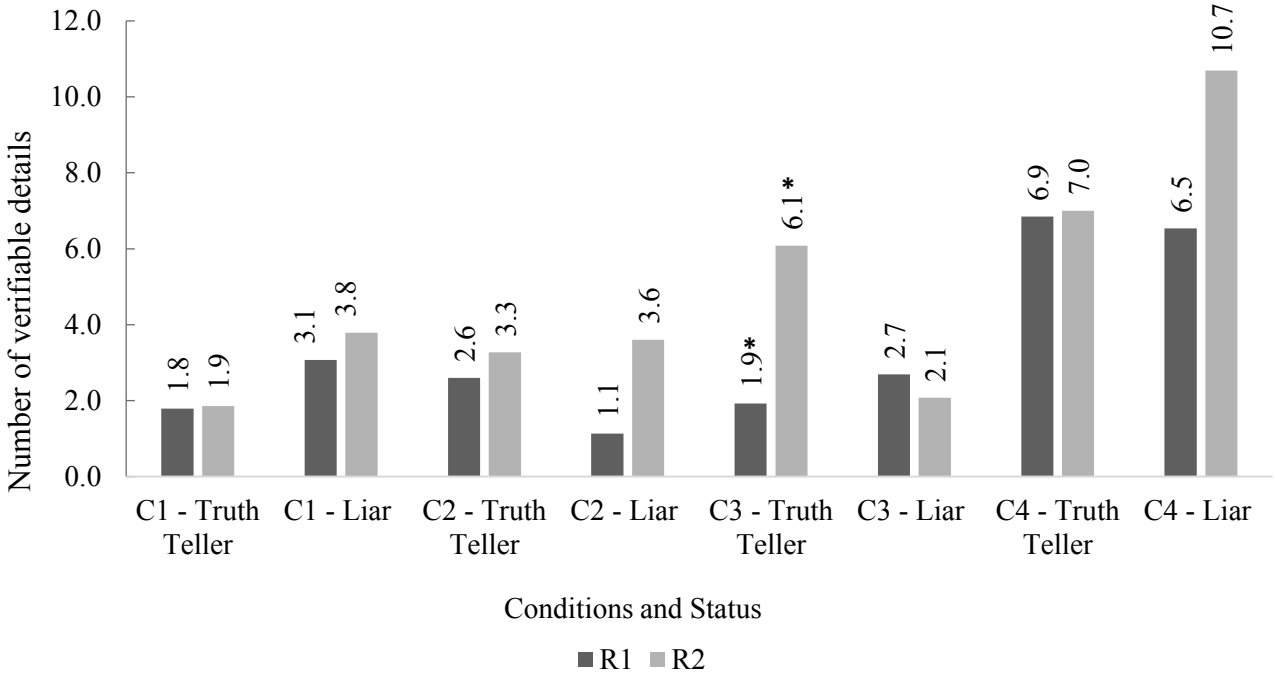
Table 4.3

3-way Mixed Effects Model for Verified Details.

	F-value	p-value
Status	0.05	.826
Recall number	7.09	.009
Condition	3.76	.013
Status*recall number	0.14	.706
Status*condition	0.41	.745
Recall number*condition	0.47	.703
Status*recall number*condition	2.72	.048

Figure 4.4

Comparison of the Number of Verifiable Details of Liars and Truth Tellers in Recalls 1 and 2 in Each Condition.



***p* = .031.**

Note: C1 = Two Basic recalls, C2 = Basic recall + Open depth recall, C3 = Basic recall + Verifiability Approach & Information Protocol, C4 = Two Verifiability Approach & Information Protocol recalls.

In contrast, there was no significant interaction between status and the recall number on the total number of verifiable details for Condition 1 (two Basic recalls), $F(1, 26) = .41$, $p = .53$, partial $\eta^2 = .016$; Condition 2 (Basic recall + Open depth), $F(1, 28) = 1.50$, $p = .23$, partial $\eta^2 = .051$, or Condition 4 (two VA & IP recalls), $F(1, 24) = 1.13$, $p = .30$, partial $\eta^2 = .045$. Detailed statistics are provided in Table 4.2. The main effect of recall number showed a significant difference in the number of verifiable details in the different recalls only for Condition 2 (Basic recall + Open depth), $F(1, 28) = 4.53$, $p = .042$, partial $\eta^2 = .14$ (large effect size), with participants giving significantly more verifiable details in R2 ($M = 3.43$, $SD = 5.00$) than in R1 ($M = 1.87$, $SD = 3.73$). There was no significant difference in the number

of verifiable details in the different recalls in Condition 1 (two Basic recalls), $F(1, 26) = .62$, $p = .44$, partial $\eta^2 = .023$, or Condition 4 (two VA & IP recalls), $F(1, 24) = 1.31$, $p = .26$, partial $\eta^2 = .052$. There was no main effect of status on the mean number of verifiable details in Condition 1 (two Basic recalls), $F(1, 26) = .88$, $p = .36$, partial $\eta^2 = .033$; Condition 2 (Basic recall + Open depth), $F(1, 28) = .15$, $p = .70$, partial $\eta^2 = .005$, or Condition 4 (two VA & IP), $F(1, 24) = .15$, $p = .70$, partial $\eta^2 = .006$.

Discriminating participants based on the evolution of verifiable details between recalls

Finally, we conducted a post hoc analysis to explore the potential for a simple within-subject decision rule to discriminate between liars and truth tellers by analysing the evolution of the number of verifiable details between a first narrative, with the Basic recall instruction, and a second narrative, with the VA & IP instruction. Specifically, we analysed the number of participants within Condition 3 (Basic recall + VA & IP) who could be correctly classified using the following decision rule: 'if a participant provides at least one more verifiable detail in their second narrative compared to their first narrative, then classify them as truthful; otherwise classify them as lying'. Based on this decision rule, 20 out of the total 26 participants in this condition (76.9%) could be correctly classified, with an equal lie and truth accuracy rate (10 out of 13 participants correctly classified in each case).

Discussion

In the present experiment, we sought to answer a recurring but under-addressed demand from practitioners for within-subject methods to evaluate the credibility of a suspect. Put differently, practitioners need methods to detect deception, taking into account individual differences in behaviors between suspects. The vast majority of research on lie detection, however, has provided general between-subject results about how liars and truth tellers may differ. For example, a general well-documented result from the literature is that liars tend to provide, on average, fewer details than truth tellers when recounting a past event (Vrij, 2008).

Interestingly, in no condition was there a significant difference in the total number of details between liars and truth tellers during the first narrative alone, therefore not replicating results from the literature on the discriminating power of either all details taken together, or verifiable details only. This result thus did not confirm Hypothesis 1 that truth tellers will provide significantly more details (both verifiable and total details) than liars during the first recall in each condition. Notwithstanding the above, it is worth noting that when experimental conditions were collapsed, the total number of details was significantly higher for truth tellers than liars in the first narrative, and more so in the second narrative, supporting hypothesis 2 that across recall instructions, truth tellers will provide significantly more details than liars, regardless of the recall number. Furthermore, this difference approached significance in the first narrative of both Conditions 1 and 2, but not Condition 3, despite all three conditions using the same basic instruction for the first recall. The lack of difference in verifiable details between liars and truth tellers is particularly surprising for Condition 4 (VA & IP instruction) as it fails to replicate findings that liars can be discriminated from truth tellers when asked for verifiable details (Harvey et al., 2017; Jupe et al., 2017; Nahari & Vrij, 2014; Vrij et al., 2016). One potential reason for this discrepancy is the medium through which the statement is provided. In their meta-analysis, Palena et al. (2021) found the effect of VA was greater for written than oral statements, possibly because written statements allow for a better organization of thought and thus allow more details to be remembered.

Regardless, knowledge of the average levels of recall for specific events is hardly transposable in the field, in the absence of clear thresholds to consider a narrative as sufficiently or insufficiently detailed to be credible. To address this issue, we have proposed and tested an interviewing method based on asking two consecutive narratives of a suspect's past activities and comparing the evolution of both the total number of details and verifiable details they provided between the first and second narrative, rather than trying to detect

deception from a single narrative only. Four sets of instructions for the first and second recall were compared. The first set of instruction served as a control condition, where the interviewees were simply asked twice to recount their activities in as much details as possible (basic recall instruction). Because it has proven to be effective in increasing the number of details provided by witnesses in the Cognitive Interview protocol, Condition 2 tested the combination of the basic instruction for the first recall with an Open depth instruction for the second recall. Finally, we also tested the Verifiability Approach & Information Protocol in Condition 3 (Basic recall for Recall 1, VA & IP instruction for Recall 2) and Condition 4 (VA & IP instructions for both recalls).

Comparing the evolution of the total number of details between the first and second narrative proved to be useful, but only for Condition 3 when analysing verifiable details, rather than all types of details in an undifferentiated manner. In this condition, only truth tellers provided more verifiable details when asked for a second narrative, while liars failed to increase the number of verifiable details they could provide in the second recall. Therefore, Hypothesis 3, that truth tellers who receive an Open depth instruction will provide significantly more additional total details compared to the first recall, was not supported. However, the results partially confirm Hypothesis 4, that truth tellers who receive VA & IP instruction in a second recall will provide significantly more additional verifiable details compared to the first recall, as it was only the case when recall one used basic recall instruction. In no other condition was a significant difference found between liars and truth tellers in the evolution of details/verifiable details between the narratives.

In order to analyse the practicability of the significant results found in Condition 3 (Basic recall + VA & IP) to differentiate liars from truth tellers in field, we explored the discriminating power of a simple within-subject decision rule for practitioners. More specifically, 76.9% of the participants in this condition could be correctly classified with the

following decision rule: ‘if a participant provides at least one more verifiable detail in their second narrative compared to their first narrative, then classify them as truthful; otherwise classify them as lying’. This accuracy, in line with the best classification rates reported in the literature, was thus obtained with a simple decision rule and not with complex statistical models.

After this initial success of our Basic instruction + VA & IP two recall within-subject method, more research is desirable. Future research could focus on why liars displayed an absence of difference in the number of verifiable details between the narratives, while truth tellers showed a significant increase. Future research should also try to independently replicate our method and especially our promising simple decision rule on a larger panel, to be more confident in recommending it broadly to practitioners in the field. Finally, the reason why Condition 4, with two VA & IP instructions, failed to discriminate liars from truth tellers, while Condition 3 (Basic recall + VA & IP) did, remains an empirical question that research could examine. Our assumption is that instructing interviewees with VA & IP in a second narrative only, after letting them search their memory in a more unconstrained manner with the first basic free recall instruction, maximizes access to memory for truth tellers.

General Discussion

The General Theory of Deception (GTD) is a new disruptive theory of lie production, prevention, and detection. For the first time, a single comprehensive deception theory attempts to describe, from end-to-end, when an individual is solicited for the truth: (1) how elementary deception modes come to mind and can be combined, (2) how they are mentally evaluated, both against each other and against the option of answering truthfully (or refusing to answer the solicitation), (3) how the evaluations of those options can be modeled as relying on five competing forces – the BPRED (Benefits, Punishment, Risk, Execution, Dissonance) factors, (4) how intrinsic or contextual time and mental capacity limits can lead to partial evaluations of those different options and lead to suboptimal decisions by the individual, and (5) how this dynamic process can be described in the form of an algorithm, thus allowing for testable predictions.

As part of the elaboration of GTD, I observed and recorded all the verbal and nonverbal strategies used to deceive a counterpart that I encountered during a 3-year period, which led to the listing of what I coined the "Elementary Deception Modes", or EDM. A total of 98 EDM were identified, allowing for the first time an estimate of the total (finite) number of possible ways to deceive a counterpart ($2^{98} - 1$ exactly, or approximately 10^{29} combinations).

Because each individual's evaluations of the BPRED factors are purely a matter of their subjective perception, a natural implication of the GTD is that an attempt, by an interviewer, to "artificially" manipulate those factors in an effort to change an individual's subjective perception should lead to a change in their tendency to lie. More specifically, increasing an individual's perceived benefits of telling the truth, and/or decreasing their perceived benefits of lying, and/or decreasing their perceived PRED factors of telling the truth, and/or increasing their perceived

PRED factors of lying should tend to deter people from engaging in deception in the first place. And even if the individual does engage in deception, we predicted that such manipulations should still tend to make the cues to deception more visible and detectable, compared to no manipulation of those factors. To confirm these predictions derived from the GTD, we conducted a series of three experiments manipulating different BPRED factors and evaluating the impact of those manipulations on either deception prevention and/or detection enhancement.

In our first experimental study, participants took part in an online negotiation conducted by email, acting as suppliers in a buyer/seller scenario. Manipulation of two GTD factors (Dissonance and Risk of lying) was easily operationalized by adding, or not, a sentence in the initial email the participants received. For example, manipulation of the Dissonance of lying factor was performed by simply adding a "I am contacting you today because *you have the reputation of being an honest and trustworthy supplier*" pro-social labeling sentence at the end of the email; while the Risk factor was manipulated using an anchor price of an alleged alternative offer. As predicted by the GTD, manipulation of the Dissonance and Risk factors of lying significantly increased the proportion of participants choosing to answer truthfully when asked their own purchasing price as a supplier – a highly sensitive question. The effects were the strongest when the manipulation of those two factors were combined, with 26.0% of participants providing a false purchasing price in the control no manipulation condition, compared to only 8.2% in the most important manipulation condition of combined Dissonance and Risk factor. Interestingly, though, as the Risk factor of lying was gradually increased in the experiment, the corresponding effect on the proportion of deceitful answers was not a perfectly linear decrease, illustrating the complex nature and possible interactions effects between BPRED factors during manipulation.

Our second experiment, a field study examining deception in real-life job applications, confirmed the effects of the BPRED factors manipulation, this time to both prevent deception and facilitate deception detection. In this study, job candidates applying for sales positions initially provided their free-form resumes, which served as a baseline of the level of inaccuracies before any BPRED factors manipulation, to the recruiting firm that posted the job offers. Still during the pre-interview stage, they were all then sent a customized application form that we designed to manipulate the Dissonance (asking the candidates at the beginning of the form to tick and sign a Yes/No box certifying that they will provide entirely accurate information in the document), the Punishment (by reminding them, at the end of the form, of the law article requiring candidates to provide information in good faith), and the Risk (by asking candidates to provide contact information of their previous job positions for potential background checks) factors of lying. Inaccuracies reduced from 23% in the free-form resumes to 11% using the customized application form. The customized form also acted as a way to effectively detect deception, as for the past professional experiences for which the candidates did not provide contact information to facilitate our verifications, only 39% contained no false information at all, compared to 77% when contact information was provided, thus making the absence of contact information a reliable cue to deceit in this context.

Importantly, in the customized form we also manipulated the Dissonance factor of lying by asking direct specific questions about different aspects of the candidates' academic and professional experiences. For example, a candidate may voluntarily give employment dates at the year level in their free-form resume to convey a false impression of a full 2-year experience, instead of an actual 2-month experience only (e.g., indicating "2015–2016", for a December 2015 – January 2016 actual employment period). However, the customized application form

asked for dates in a day/month/year format, forcing such candidates to either rectify this ambiguity and provide the accurate employment dates (i.e., becoming "more honest"), or engage in a real falsification lie in the customized application form (e.g., lying by reporting 01/01/2015 – 31/12/2016). Such direct close-ended or multiple-choice questions force candidates who voluntarily provided ambiguously deceptive information in their free-form resumes to now choose between rectifying any such ambiguity, and thus providing fully truthful information in the customized application form, or engage in a less morally acceptable and more mentally conflicting bald-faced lie. The use of direct and specific questions is generally not advised in the deception literature (Vrij, 2008) as they are considered detrimental to information gathering and deception detection. However, the GTD framework predicts that they increase the Dissonance factor of lying, and thus either prevent deception or increase the chances of cues to deceit appearing in response to such questions. We therefore make a fundamental distinction between questioning technique and the type of deception one is trying to address. When the aim is to assess the credibility of an account of past activities, an interviewer should start with an open-ended information gathering phase to maximize the number of details disclosed. Direct questions can then be asked based on this crucial data-gathering phase that also allows interviewers to better determine which direct questions to ask at the end of the interview. By contrast, for deception about specific facts (e.g., the salary a candidate had at their previous position), or when an interviewee is denying any wrongdoing (e.g., denying having an illegal foreign bank account), asking for open narratives might not be applicable, so deception detection can rely solely on direct questions. As an example, we can consider the case of former French budget Minister Jérôme Cahuzac. On December 4, 2012, he was accused by the journal "Mediapart" of having had an illegal bank account in Switzerland until 2010 (a fact he later acknowledged

publicly). The following day, a journalist asks Cahuzac direct questions during a radio interview, and his reactions show several potential deception modes. Typically, when asked if the Mediapart journalists "invented everything", he stated: "I do not know what their intentions are and the elements at their disposal. I would like them to show them". Such an unspecific non-Yes/No answer (here, *Deceptive Diversion EDM 77 – Topic Switch*), to a direct Yes/No question, should be considered a potential cue to deception, confirming that cues to deception can appear as a result of a direct question. The choice of such an answer by Jérôme Cahuzac, as opposed to simply falsely answering something along the lines of "Yes, these journalists are inventing everything and their allegations are false" (*Falsification EDM 15 – False Response*), is well interpreted in GTD terms. Simply put, in this case, EDM 77 probably raised a lower mental conflict (i.e., lower Dissonance factor) in Jérôme Cahuzac than EDM 15, while also being perceived by the former minister as being less reprehensible (lower Punishment factor) and having a better Justification power ("I did not state anything factually wrong in my statement") if later shown to have been lying.

Our last experimental testing of the GTD focused on deception detection enhancement, manipulating a more commonly studied factor in the deception literature: cognitive load (Vrij et al., 2008, 2009, 2012; Walczyk et al., 2005, 2009, 2012, 2013). In the GTD, cognitive load represents a subfactor of the GTD's Execution cost factor, comprising the cognitive, emotional, and control/inhibition costs. In this experiment, participants took part in a forensic-like scenario, and either chose to be truthful (conducting their regular activities) or to lie about their past activities (while they were actually committing a mock theft of a 10-euro store voucher). All participants were informed that they would be interviewed as suspects of a theft that occurred while they were near the building where the theft took place. To better simulate real-life forensic

interviewing conditions, several GTD factors were manipulated in the same manner for all participants: (1) while both liars and truth tellers had an incentive of trying to be believed by the interviewer to be granted course credits, liars perceived a greater benefit in persisting to lie during the interview as they would, if believed by the interviewer, also be able to keep the store voucher they stole after the interview, which translates in GTD terms to $B_{Lying} > B_{Truth\ telling}$; (2) like in real-life trials, whether they had committed the mock crime (liars) or not (truth tellers), all participants were informed that if the interviewer did not believe they were innocent, they would have the same punishment (in this case to write a one-page essay explaining their opinion on the sanctions that should be given to students who commit theft on campus), thus increasing GTD's Punishment factor; (3) importantly, rather than being assigned, participants chose whether they wanted to participate as a liar or a truth teller. Not only was this freedom of choice for participants (almost never allowed in other deception experiments) believed to be more realistic, it was also intended to increase the Dissonance factor in participants choosing to act as liars. In other words, we believed that letting a participant choose to be deceptive or not would engage the participants more, and make the cues to deception more salient during the interview, as opposed to externally assigning them to the lie condition. While the above manipulations of GTD factors were equal for all participants, four different sets of instructions were tested in this experiment to manipulate cognitive load in participants having to recount their past activities. Of those, only one set of instructions, asking participants to first recount their activities with a basic free recall instruction, followed by a request for a second recall asking for verifiable details as indicators of truth telling (the "Verifiability Approach & Information Protocol" instruction) proved to be useful in discriminating liars from truth tellers. This experiment therefore illustrates that manipulation of GTD factors to improve deception detection can require pilot experimental

testing of different variants of the factors' manipulation to effectively produce the intended outcome, prior to large scale real-life deployment by practitioners.

While we believe the current work and elaboration of the GTD is an important and necessary contribution to the (mostly atheoretical) field of research on deception, this disruptive theory also opens an avenue of new questions and future research to be conducted. First, future research should now explore how to derive a systematic interviewing method for credibility assessment from the GTD. We believe the following represent promising practical interviewing guidelines derived from the GTD, to be confirmed experimentally:

(1) *Defining the motive*. The GTD posits that when an explicit or implicit solicitation of truth is made to an individual, this automatically triggers in them an initial mental evaluation of whether *not disclosing* the requested information can have any benefit to them. If that is not the case ($B_{ND} = 0$), then the GTD assumes the individual chooses to answer truthfully. Simply put, people do not consider lying unless they perceive some benefit to it. Therefore, as is already a common practice in police professionals but rarely put forward in the deception literature, when trying to assess credibility of an individual, one should probably always start by evaluating what interest the person questioned could have in lying. For example, if a public figure who is running for an important position is accused of historical sexual abuse by a successful and respected professional with no financial difficulties, which the accused strongly denies, then considering who benefits from lying is of paramount importance in the overall credibility assessment of the accuser. In this case the alleged victim does not seem to benefit from lying whereas the alleged perpetrator certainly would. In other words, analysing the potential motives of the different stakeholders already

provides some contextual elements of credibility assessment, independently of any specific statement analysis from the protagonists.

(2) *Evaluating the Base Rate*. Much in the same vein, proper assessment of the credibility of a statement can also benefit from evaluating how prevalent the kind of claim described is, as a contextual element of the case analysis, independently of the statement's potential cues to deception. In other words, if someone describes something that very rarely happens in everyday life, this Base Rate should be weighted in the overall credibility assessment of the case.

(3) *Fostering honesty. Facilitating lie detection*. As has been partially examined and demonstrated in the present work and experimental studies testing predictions derived from the GTD and the Five Forces Model, interviewers should start their interview by manipulating the BPRED factors as much as possible in an attempt to make people more honest and deter deception from happening in the first place. More specifically, the GTD predicts that the likelihood that an individual will engage in deception can be manipulated by any combination of the following: increasing the perceived benefits of telling the truth, and/or decreasing the perceived benefits of lying, and/or decreasing the PRED factors of telling the truth, and/or increasing the PRED factors of lying. Examples include mentioning that making 100% honest statements, whatever their content, has always led to more favorable outcomes (i.e., increasing B_T); making the interviewee aware that their interviewer is a recognized expert in credibility assessment (increasing R_L); building rapport or using pro-social labels such as "you seem to be an honest and trustworthy person" (increasing D_L); or discussing, in a non-accusatory manner, the consequences of being found guilty

(increasing P_L), e.g., by stating "As part of this procedure, I must also inform you that as a general rule, undeclared work can be subject to penalties of up to 45,000 EUR, a ban on working in the profession, criminal penalties up to 3 years of imprisonment and a publication of the judgment". Such manipulations are also meant to make the cues to deception more salient in interviewees still engaging in deception, thus facilitating residual lie detection during the interview. While the literature to date has mainly focused on manipulating cognitive load (E factor in the GTD) in interviews to better detect deception (Vrij et al., 2008, 2009, 2012; Walczyk et al., 2005, 2009, 2012, 2013), much remains to be tested regarding the other BPRED factors' manipulations and their effect on deception detection enhancement.

Finally, careful attention should be made in the manipulation of the BPRED factors to avoid reverse effects. For example, bluffing a suspect by saying "look, we already have all the evidence against you, just confess to the murder so you can avoid a life sentence", with the intention of deterring suspects from lying by increasing the P_L punishment factor, can have devastating effects on a wrongfully accused innocent suspect hearing such a statement (i.e., increasing P_T in truth tellers and potentially leading to false confessions). Each manipulation of a BPRED factor should therefore be well thought and designed to work and be acceptable for both liars and truth tellers.

(4) *Free recalls and reminiscence boosting instructions*. When trying to assess the credibility of past activities (e.g., was a person of interest actually at the office working last Tuesday, between 1 and 3pm), interviews, as already largely recommended in the deception literature (Vrij, 2008), should start by asking the

interviewees for open narratives in the form of unguided description of their past activities. Such "free recalls", in the GTD's description of memory processes, allow truth tellers to activate and make accessible in memory as many relevant specific expected items as possible. As a reminder, in the GTD, there is only one truthful answer in response to a solicitation for truth, composed of all the relevant specific expected items (and only those) activated in memory at time t_{end}. Letting enough time for a truthful individual to remember a maximum of information, or even using free recall instructions designed to boost reminiscence of past events, thus lead to more activations of relevant specific expected items in memory, and more complete descriptions of events than truth tellers interviewed without the use of such instructions. While one could argue that allowing more time for truth tellers to remember also gives more time for liars to elaborate their story, we believe that this is only true for a short time, beyond which liars, as documented in the literature (Strömwall et al., 2006), tend to want to keep their story simple (lowering their Execution cost factor) and unverifiable (lowering the probability $prob(undesirable\ event)$ of later undesirable events, such as contradiction with factual elements). On a practical side, initial free recalls also allow interviewers to discover areas of suspicion in a testimony that they were not previously aware of.

(5) *"Nobody wants to lie". A new deception detection principle*. In the GTD's Deception Decision Algorithm (DDA), the different options one mentally compares in response to a solicitation for truth are modeled using Utility functions. The option with the highest score is the one selected for answering the solicitation. Between the deceptive options mentally considered, we posit that the ones the further from Falsification

deception modes (from the EDM categories defined in the GTD, i.e., Deceptive Omission, Falsification, Misinterpretation, Distortion, Deceptive Evasion, and Deceptive Diversion), will generally have the highest Utility functions scores. While this statement of course depends on the context and the exact way the questions are framed to the interviewee, most of the time, non-Falsification deception modes should be preferred as they lower the Execution cost factor (it is less cognitively and emotionally demanding to omit a detail than to fabricate one, for example), the Dissonance factor (falsifying an answer is more mentally and morally conflicting than dodging a question) and the Punishment factor (people believe they will have less severe punishments if they think they "did not really lie", as could be the case with non-Falsification deception modes), and for the same reason, increase the perceived Justification power of such deception modes ("I can always say I did not really lie"). For all these reasons, the GTD predicts that the Five Forces push people to opt for deception that is the closest to what they believe is the truth. Whereas coding for all 98 potential EDM in a statement as a means to detect deception, can prove impractical in real life due to the high number of EDM to code, we propose reducing deception detection to one basic principle derived from the GTD. *Nobody wants to lie*. This does not mean that nobody lies. Rather, this means that people, when they do not want to disclose the truth, try as much as possible to be deceptive while keeping the conscious or unconscious impression they are not really lying. What my colleagues have now called the "Srour principle" can indeed, I believe, be a very simple and efficient way to better detect deception, especially in response to direct questions, which I will illustrate below.

(6) *Defining the suspicion. Closing with Direct questions*. *Detecting direct traces of verbal non-Falsification deception*. Because I posit that *Nobody wants to lie* and that people tend to try to minimize their deviation from the truth when answering deceitfully, as an interviewer trying to assess credibility, defining the exact suspicion of deception one has at the beginning of the interview is of paramount importance. It is also a necessary prerequisite to apply the *Nobody wants to lie* principle as a means to detect lies. As a simple example (the one chosen to illustrate *EDM 68 – Implied Response*), imagine a police officer asking a suspect: "Did you kill your wife?". The latter responds: "I have never hurt anybody!". Recognizing that the police officer's suspicion is whether the suspect killed his wife, and applying the *Nobody wants to lie* principle, the police officer should immediately notice that the suspect could actually have killed his wife (the suspicion) while thinking he did not make her suffer in the process, making the suspect (unconsciously?) believe he technically did not lie, and probably reducing his Dissonance factor with such an answer. Using the Srour principle with a clearly defined suspicion (which is not always possible; for example one cannot ask a car seller if a specific part of the vehicle has to be replaced soon if they do not even suspect that kind of scam from the seller) can thus, we believe, allow practitioners to detect traces of all non-Falsification deception. As this principle increases the detection of any deviation or ambiguity in response to direct questions, we thus promote again here the use of such questions to directly address the interviewer's suspicion, at the end of the interview, when they can no longer "pollute" reminiscence or rapport building with the interviewee. Combining the Srour principle with direct final questions addressing the suspicion at stake thus forces the

deceptive interviewee to choose one of three options: (i) reiterating non-Falsification lies, which as described above leave direct detectable traces (i.e., not answering a specific direct question with a perfectly specific answer), making the probability of deception from the interviewee even higher, (ii) opting for a Falsification lie, (iii) getting away from the previous ambiguity of non-Falsification lies, and telling the truth. For example, going back to the example used to illustrate *EDM 1 – Narrative Omission*, imagine a husband told his wife he would be home from work early today, but ends up arriving more than an hour and a half late. When his wife asks him what happened, he answers: "I had a lot of work today". Even though he indeed had an intense day at work, he fails to mention he still found some time to have a drink with his colleagues after work before heading back home. If, in similar situations in the past, his wife had discovered that he had "wasted" some time with colleagues before heading back home, she could now have the same suspicion whenever he was home late (i.e., she has a clearly identified suspicion). She therefore asks him this time: "Did you come straight home after work?" (direct question addressing the suspicion). The husband is now forced to either persist in non-Falsification deception ("Honey, you know how hard I have been working lately, let's not waste any more time and start our evening"), admit to the truth ("No, I had some beers with colleagues after work first") or engage in a Falsification lie ("Yes, I came straight home after work").

(7) *Detecting Falsification and Nonverbal lies*. *Using W questions. Detecting indirect traces of deception*. Following the interviewing method we suggested should thus lead to either easily detectable lies, people telling the truth, or falsification (and all categories of nonverbal) lies that leave no direct traces detectable with the Srour

principle. What to do in such cases? Firstly, verbal falsification lies can be roughly divided into two types: those involving a fabrication of a narrative (e.g., inventing activities as an alibi), and those simply involving the statement of factually incorrect information (e.g., replacing a truthful "No" with a false "Yes" in response to a question, see example above at the end of point (5)). While fabricated narratives can technically be addressed with direct questions as well, quite often, the suspicion is not well-defined enough for relevant direct questions to be asked. If for example a youngster is suspected of inventing a story about his bus having mechanical issues to explain to his professor why he is more than 20 minutes late, the professor could of course ask the direct question "Are you lying to me now?", but this would be very unspecific to the suspicion of fabrication and might not prove useful. Instead, one should consider that a narrative fabrication *always* involves at least one of the following fabrications: fabrication of time (e.g., a true story that happened at another time), and/or fabrication of location (e.g., a true story that happened elsewhere), and/or fabrication of protagonists (e.g., a true story that happened to other people), and/or fabrication of the core event (e.g., an invented event in truly existing locations, with protagonists truly present at the time of the alleged event). Therefore, after asking for a free recall of the event, an interviewer assessing the credibility of the narration should always ask for more details with all the "W questions" around the event: When, Where, Who, What? Such requests for more details will at some point specifically target the fabrication that was performed by a liar, thus increasing the chances of making *indirect traces of deception* (e.g., lack of details, deviations of nonverbal behaviors vs comparable questions during the same interview, longer

response times, etc.) more salient, as opposed to the direct traces of verbal non-Falsification deception detectable with the Srour principle in response to direct questions addressing the suspicion. Second, the detection of false factual statements (replacing a Yes with a No in response to a question, giving a false purchasing price, etc.), as well as detecting nonverbal deception, only leaves the interviewer with the analysis of nonverbal cues as potentially revealing deception. While manipulation of the BPRED factors (see point (3)) could theoretically help magnify such cues, there is to date no consensus in the scientific community, or published compelling evidence, that there is a simple and reliable way of detecting deception from nonverbal cues. Therefore, how should an interviewer proceed when they suspect lies with false facts?

(8) *External and Internal Consistency Tests*. Detection of Falsification lies thus appear to be the hardest to detect, especially the ones simply stating factually incorrect information. Fortunately, as explained above (point (5)), the GTD predicts that those lies should be the less prevalent ones. Yet, they can also be the most important to detect. To address this specificity, we propose the credibility assessment interviews to include what I call "External and Internal Consistency Tests". Simply put, the interviewer should, after the free recall and before the direct questions phases, orientate and asks questions whose answers they can compare against (a) external facts or evidence (External Consistency Test), or (b) other previous answers from the interviewee (Internal Consistency Test). Here is an example of an External Consistency Test. If someone claims to have finished a regional marathon in 10th place, the interviewer should orientate their questions to get a maximum of verifiable information about the event, and then actually check (if any) the published marathon

ranking on the internet. Although both asking for verifiable details and checking whether the interviewee has provided verifiable details have already been suggested and successfully tested as a cue of truth telling in the literature (Harvey et al., 2017; Jupe et al., 2017; Nahari & Vrij, 2014; Vrij et al., 2016), here, we go a step further. Simply put, we state that *what is better than a verifiable detail, is a verified detail*. In other words, we recommended not only checking whether an interviewee has provided verifiable details, but trying as much as possible to orientate the interview questions to have interviewees state information that can be verified, try to verify it, and confirm or contradict the interviewee's statement. Then, if the nature or context of a statement justifies the absence of external corroborating evidence (e.g., someone talking about an incident that happened at work during a confidential business meeting), the interviewer should resort to Internal Consistency Tests. For example, the interviewer could first ask the interviewee to verbally describe, in as much detail as possible, who was seated where around the negotiation table, and much later in the interview, to sketch the configuration and protagonists around the table, then compare the consistency of the answers. Conducting such tests during the interview, especially External Consistency Tests, also have the benefit, in GTD terms, of increasing the deceitful interviewee's perception of "the likelihood of an undesirable event" of lying (i.e., that their version will be checked and might be contradicted by external evidence), thus decreasing the Utility functions scores of their deceptive options.

(9) *Fostering confessions.* Finally, much like we suggested the BPRED factors should be manipulated at the beginning of the interview to deter people from engaging in deception, we suggest such manipulation would also be beneficial at the very end of

the interview to foster confessions from liars. Again, this manipulation should be carefully designed to avoid any undesired effect on truth tellers. Here is an example of such a manipulation that could be conducted by an investigator interviewing a restaurant owner about a suspicion of undeclared labor. After first asking what reasons could explain, in their opinion, *other* restaurant owners resorting to undeclared labor (to which the interviewed restaurant owner answers "maybe because they think the taxes are just too high to make ends meet"), the investigator replies with the following: "You are right, my colleague and I do indeed meet, through our profession, many restaurant owners who mention the reasons you have cited. It's very common. *When they tell us about it, we do our best to help them regularize their situation, whatever their previous claims were during our investigation*". In this example, the manipulation is thus designed to increase the perceived benefit of telling the truth ("we do our best to help them regularize their situation"), while reducing the dissonance ("whatever their previous claims was during our investigation") they may feel of confessing to undeclared labor after having denied it repeatedly so far.

Besides the nine practical interviewing guidelines outlined above, other axes of research derived from the GTD should be explored. Empirical questions include whether the guidelines derived from the GTD can (1) lead to better credibility assessment, and (2) be effectively taught to professionals. Our preliminary data on these questions, comparing pre- and post-training lie detection accuracies of professional investigators, suggest a +25 points improvement in accuracy after a 4-day training program – but much more data is needed and continues to be collected to confirm those pilot results.

Adapting the interviewing framework to different types of deception also represent a crucial theoretical question. While the guidelines outlined above can all be applied during an interview aiming to detect lies about *past activities*, the remaining types of lies might not be apprehended with all the steps suggested above, and may even require new interviewing strategies. Namely, we consider that deception can be divided into the following six types of lie:

(1) *Lies about past activities* (e.g., a person lying about their activities in response to the question "What were you doing last Saturday between 8 and 11pm?)

(2) *Lies about facts* (e.g., a person lying in response to "What is your current salary?")

(3) *Deceitful denials* (e.g., a person lying in response to "Do you, or did you at any time, have a bank account in Switzerland?")

(4) *Lies about future activities* (e.g., a person lying in response to "What are you planning to do during your stay in New York City?")

(5) *Lies about opinions* (e.g., a person lying in response to "What do you think of this ideological group?")

(6) *Lies about motivations* (e.g., a person lying in response to "What is the reason of your stay in the country?")

Finally, future experiments on the GTD should address basic research questions related to the theory, like the strength of potential interactions between BPRED factors (e.g., does decreasing the Dissonance factor also systematically decreases the emotional component of the Execution cost factor; are some BPRED factors more correlated than others, etc.), the non-linear manipulation of BPRED factors (e.g., threatening a candidate that any discovered deception in their application would be registered in a shared recruiter database might be a deterrent, while threatening resume fraud with the death penalty would of course not be deterrent at all as it

would not be believed by any candidate), the interest of splitting the concerned BPRED factors into their subfactors (e.g., splitting the Execution cost factor into cognitive, emotional, and control/inhibition costs factors), and looking for traces of item activations, BPRED evaluation, and DDA processes through neuro-imaging techniques.

After an historic trend in the deception research community for passive deception detection, followed by active deception detection using cognitive interviewing strategies to magnify cues to deception, the GTD now paves a new way for deception *prevention*. More generally, from basic research into real-life applications in professional interviews, the GTD opens a myriad of new and exciting studies to be conducted in the next decade.

References

Aquino, K. (1998). The effects of ethical climate and the availability of alternatives on the use of deception during negotiations. *International Journal of Conflict Management*, *9*(3), 195–217. https://doi.org/10.1108/eb022809

Aquino, K., & Becker, T. E. (2005). Lying in negotiations: How individual and situational factors influence the use of neutralization strategies. *Journal of Organizational Behavior*, *26*(6), 661–679. https://doi.org/10.1002/job.332

Bazerman, M. H., Curhan, J. R., Moore, D. A., & Valley, K. L. (2000). *Negotiation. Annual Review of Psychology, 51*, 279–314. https://doi.org/10.1146/annurev.psych.51.1.279

Bazerman, M. H., Giuliano, T., & Appelman, A. (1984). Escalation of commitment in individual and group decision making. *Organizational Behavior and Human Performance*, *33*(2), 141–152. https://doi.org/10.1016/0030-5073(84)90017-5

Beggs, A., & Graddy, K. (2009). Anchoring Effects: Evidence from Art Auctions. *American Economic Review*, *99*(3), 1027–1039. https://doi.org/10.1257/aer.99.3.1027

Belot, M., & van de Ven, J. (2017). How private is private information? The ability to spot deception in an economic game. *Experimental Economics*, *20*(1), 19–43. https://doi.org/10.1007/s10683-015-9474-8

Bendersky, C., & Curhan, J. R. (2009). Cognitive dissonance in negotiation: Free choice or justification? *Social Cognition*, *27*(3), 455–474. https://doi.org/10.1521/soco.2009.27.3.455

Bokhari, S., & Geltner, D. (2011). Loss Aversion and Anchoring in Commercial Real Estate Pricing: Empirical Evidence and Price Index Implications. *Real Estate Economics*, *39*(4), 635–670. https://doi.org/10.1111/j.1540-6229.2011.00308.x

Boles, T. L., Croson, R. T. A., & Murnighan, J. K. (2000). Deception and Retribution in Repeated Ultimatum Bargaining. *Organizational Behavior and Human Decision Processes*, *83*(2), 235–259. https://doi.org/10.1006/obhd.2000.2908

Bond, C. F., & DePaulo, B. M. (2006). Accuracy of Deception Judgments. *Personality and Social Psychology Review*, *10*(3), 214–234. https://doi.org/10.1207/s15327957pspr1003_2

Bourdage, J. S., Roulin, N., & Tarraf, R. (2018). 'I (might be) just that good': Honest and deceptive impression management in employment interviews, *Personnel Psychology*, *71*(4), 597–632. https://doi.org/10.1111/peps.12285

Brunel, M., Py, J., & Launay, C. (2013). Cost and benefit of a new instruction for the cognitive interview: The open depth instruction. *Psychology, Crime & Law*, *19*(10), 845–863. https://doi.org/10.1080/1068316X.2012.684058

Bull, R. (2010). The investigative interviewing of children and other vulnerable witnesses: Psychological research and working/professional practice. *Legal and Criminological Psychology*, *15*(1), 5–23. https://doi.org/10.1348/014466509X440160

Buller, D. B., Burgoon, J. K., Buslig, A., & Roiger, J. (1996). Testing Interpersonal Deception Theory: The Language of Interpersonal Deception. *Communication Theory*, *6*(3), 268–289. https://doi.org/10.1111/j.1468-2885.1996.tb00129.x

Buller, D. B., Burgoon, J. K., White, C. H., & Ebesu, A. S. (1994). Interpersonal deception: VII. Behavioral profiles of falsification, equivocation, and concealment. *Journal of Language and Social Psychology*, *13*(4), 366–395. https://doi.org/10.1177/0261927X94134002

Carr, A. Z. (1968, January 1). Is Business Bluffing Ethical? *Harvard Business Review*, *January 1968*. https://hbr.org/1968/01/is-business-bluffing-ethical

Chandrashekaran, R., & Grewal, D. (2006). Anchoring effects of advertised reference price and sale price: The moderating role of saving presentation format. *Journal of Business Research*, *59*(10), 1063–1071. https://doi.org/10.1016/j.jbusres.2006.06.006

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Routledge.

Clifford, B. R., & George, R. (1996). A Field evaluation of training in three methods of witness/victim investigative interviewing. *Psychology, Crime & Law*, *2*(3), 231–248. https://doi.org/10.1080/10683169608409780

Collins, R., Lincoln, R., & Frank, M. G. (2002). The Effect of Rapport in Forensic Interviewing. *Psychiatry, Psychology and Law*, *9*(1), 69–78. https://doi.org/10.1375/pplt.2002.9.1.69

Colwell, K., Hiscock-Anisman, C., & Fede, J. (2013). Assessment Criteria Indicative of Deception: An Example of the New Paradigm of Differential Recall Enhancement. In B. S. Cooper, D. Griesel, & M. Ternes (Eds.), *Applied Issues in Investigative Interviewing, Eyewitness Memory, and Credibility Assessment* (pp. 259–291). Springer. https://doi.org/10.1007/978-1-4614-5547-9_11

Colwell, K., Hiscock-Anisman, C. K., Memon, A., Taylor, L., & Prewett, J. (2007). Assessment Criteria Indicative of Deception (ACID): An integrated system of investigative interviewing and detecting deception. *Journal of Investigative Psychology and Offender Profiling*, *4*(3), 167–180. https://doi.org/10.1002/jip.73

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin*, *104*(2), 163–191. https://doi.org/10.1037/0033-2909.104.2.163

Cowan, N. (1998). *Attention and Memory: An Integrated Framework*. Oxford University Press. doi:10.1093/acprof:oso/9780195119107.001.0001

Cowan, N. (1999). *An Embedded-Processes Model of Working Memory*. Models of Working Memory: Mechanisms of Active Maintenance and Executive Control. https://doi.org/10.1017/CBO9781139174909.006

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24*(4), 349–354. https://doi.org/10.1037/h0047358

DePaulo, B. M. (1992). Nonverbal behavior and self-presentation. *Psychological Bulletin*, *111*(2), 203–243. https://doi.org/10.1037/0033-2909.111.2.203

DePaulo, B. M., & Kirkendol, S. E. (1989). The Motivational Impairment Effect in the Communication of Deception. In J. C. Yuille (Ed.), *Credibility Assessment* (pp. 51–70). Springer Netherlands. https://doi.org/10.1007/978-94-015-7856-1_3

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. https://doi.org/10.1037/0033-2909.129.1.74

Drummond, H. (1994). Escalation in organizational decision making a case of recruiting an incompetent employee. *Journal of Behavioral Decision Making*, *7*(1), 43–55. https://doi.org/10.1002/bdm.3960070104

Dukala, K., Sporer, S. L., & Polczyk, R. (2019). Detecting deception: Does the cognitive interview impair discrimination with CBCA criteria in elderly witnesses? *Psychology, Crime & Law*, *25*(2), 195–217. https://doi.org/10.1080/1068316X.2018.1511789

Dwight, S. A., & Donovan, J. J. (2003). Do warnings not to fake reduce faking? *Human Performance, 16*(1), 1–23. https://doi.org/10.1207/S15327043HUP1601_1

Eklinder-Frick, J., & Åge, L.-J. (2016). *Transactional and interactional perspectives on business negotiation*. 21. http://urn.kb.se/resolve?urn=urn:nbn:se:hig:diva-23139

Ekman, P. (1985). *Telling Lies.* W. W. Norton & Company.

Ekman, P. (1991). *Why Kids Lie*. Penguin.

Ekman, P. (2009). *Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage (Revised Edition)*. W. W. Norton & Company.

Ekman, P. (2016). *Nonverbal Messages: Cracking the Code: My Life's Pursuit*. Paul Ekman Group.

Ekman, P., & Friesen, W. V. (1969). Nonverbal Leakage and Clues to Deception. *Psychiatry*, *32*(1), 88–106. https://doi.org/10.1080/00332747.1969.11023575

Ekman, P., & O'Sullivan, M. (1991). Who can catch a liar?. The American psychologist, *46*(9), 913–920. https://doi.org/10.1037//0003-066x.46.9.913

Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Stanford University Press, California.

Fisher, R., Ury, W. L., & Patton, B. (2011). *Getting to yes: Negotiating agreement without giving in* (Updated, Revised). Penguin Books.

Fisher, R. P. (2010). Interviewing cooperative witnesses. *Legal and Criminological Psychology*, *15*(1), 25–38. https://doi.org/10.1348/135532509X441891

Fisher, R. P., & Geiselman, R. E. (1992). *Memory-enhancing techniques for investigative interviewing: The cognitive interview* (pp. xi, 220). Charles C Thomas, Publisher.

Florian Mantione Institut. (2017), "8e étude sur les CV Trompeurs [8th Study on Deceptive Resumes]", Retrieved from https://www.florianmantione.com/images/files/CVTrompeurs_Etude_8.pdf

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.

Gasper J. P., & Schweitzer, M. E. (2013). The emotion deception model: A review of deception in negotiation and the role of emotion in deception. *Negotiation and Conflict Management Research*, *6*(3), 160–179. https://doi.org/10.1111/ncmr.12010

Geiger, S. W., Irwin, J. G., & Robertson, C. J. (1998). The impact of cultural values on escalation of commitment. *The International Journal of Organizational Analysis*, *6*(2), 165–176. https://doi.org/10.1108/eb028883

Geiselman, R. E. (2012). The cognitive interview for suspects (CIS). *American Journal of Forensic Psychology*, *30*(3), 5–20.

Geiselman, R. E., Fisher, R. P., Firstenberg, I., Hutton, L. A., Sullivan, S. J., Avetissian, I. V., & Prosk, A. L. (1984). Enhancement of eyewitness memory: An empirical evaluation of the cognitive interview. *Journal of Police Science & Administration*, *12*(1), 74–80.

Geiselman, R. E., Fisher, R. P., MacKinnon, D. P., & Holland, H. L. (1986). Enhancement of Eyewitness Memory with the Cognitive Interview. *The American Journal of Psychology*, *99*(3), 385–401. JSTOR. https://doi.org/10.2307/1422492

Ginet, M., & Py, J. (2001). A technique for enhancing memory in eye witness testimonies for use by police officers and judicial officials: The cognitive interview. *Le travail humain*, *Vol. 64*(2), 173–191.

Gino, F., & Shea, C. (2012). Deception in Negotiations: The Role of Emotions. In R. Croson & G. E. Bolton (Eds.), *The Oxford Handbook of Economic Conflict Resolution* (pp. 46–60). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199730858.013.0005

Giordano, G. A., Stoner, J. S., Brouer, R. L., & George, J. F. (2007). The Influences of Deception and Computer-Mediation on Dyadic Negotiations. *Journal of Computer-Mediated Communication*, *12*(2), 362–383. https://doi.org/10.1111/j.1083-6101.2007.00329.x

Gödert, H. W., Gamer, M., Rill, H.-G., & Vossel, G. (2005). Statement validity assessment: Inter-rater reliability of criteria-based content analysis in the mock-crime paradigm, *Legal and Criminological Psychology*, *10*(2), 225–245. https://doi.org/10.1348/135532505X52680

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. https://doi.org/ 10.3200/MONO.132.3.197-214

Grice, H. P. (1989). *Studies in the Way of Words*. Harvard University Press. http://www.hup.harvard.edu/catalog.php?isbn=9780674852716

Gumpert, C. H., & Lindblad, F. (2000). Expert testimony on child sexual abuse: A qualitative study of the Swedish approach to statement analysis. *Expert Evidence*, *7*(4), 279–314. https://doi.org/10.1023/A:1016657130623

Hample, D. (1980). Purposes and effects of lying. *Southern Speech Communication Journal*, *46*(1), 33–47. https://doi.org/10.1080/10417948009372474

Harmon-Jones, E., Amodio, D. M., & Harmon-Jones, C. (2009). Action-based model of dissonance: A review, integration, and expansion of conceptions of cognitive conflict. In *Advances in experimental social psychology, Vol 41* (pp. 119–166). Elsevier Academic Press.

Harvey, A. C., Vrij, A., Nahari, G., & Ludwig, K. (2017). Applying the Verifiability Approach to insurance claims settings: Exploring the effect of the information protocol. *Legal and Criminological Psychology*, *22*(1), 47–59. https://doi.org/10.1111/lcrp.12092

Hayes, L. N. (2017). *75% of HR Managers have caught a lie on a resume, according to a new CareerBuilder survey*. Retrieved from http://press.careerbuilder.com/2017-09-14-75-of-HR-Managers-Have-Caught-a-Lie-on-a-Resume-According-to-a-New-CareerBuilder-Survey

Henle, C. A., Dineen, B. R., & Duffy, M. K. (2019). Assessing intentional resume deception: Development and nomological network of a resume fraud measure. *Journal of Business and Psychology, 34*, 87–106. https://doi.org/10.1007/s10869-017-9527-4

Ingold, P. V., Kleinmann, M., König, C. J., & Melchers, K. G. (2015). Shall we continue or stop disapproving of self-presentation? Evidence on impression management and faking in a selection context and their relation to job performance. *European Journal of Work and Organizational Psychology, 24*, 420–432. https://doi.org/10.1080/1359432X.2014.915215

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review*, *88*(1), 67–85. https://doi.org/10.1037/0033-295X.88.1.67

Jupe, L., Vrij, A., Leal, S., Mann, S., & Nahari, G. (2016). The Lies we Live: Using the Verifiability Approach to Detect Lying about Occupation. *Journal of Articles in Support of the Null Hypothesis*, *13*(1), 14.

Jupe, L. M., Leal, S., Vrij, A., & Nahari, G. (2017). Applying the Verifiability Approach in an international airport setting. *Psychology, Crime & Law*, *23*(8), 812–825. https://doi.org/10.1080/1068316X.2017.1327584

Kersten, G. E., Köszegi, S. T., & Vetschera, R. (2002). The effects of culture in anonymous negotiations: Experiment in four countries. *Proceedings of the 35th Annual Hawaii*

International Conference on System Sciences. https://doi.org/10.1109/HICSS.2002.993890

Köhnken, G., & Steller, M. (1988). The evaluation of the credibility of child witness statements in the German procedural system. *Issues in Criminological & Legal Psychology*, *13*, 37–45.

Lamers-Wilkelman, F. (1999). Statement validity analysis: Its application to a sample of Dutch children who may have been sexually abused. *Journal of Aggression, Maltreatment & Trauma*, *2*(2), 59–81. https://doi.org/10.1300/J146v02n02_04

Le Paige, H. (Director). (2011). *Le Prince et son Image.* [Film]. Arturo Mio.

Levashina, J., & Campion, M. A. (2009). Expected Practices in Background Checking: Review of the Human Resource Management Literature. *Employee Responsibilities and Rights Journal*, *21*(3), 231–249. https://doi.org/10.1007/s10672-009-9111-9

Levashina, J., Morgeson, F.P., & Campion, M. A. (2012). Tell me some more: Exploring how item verifiability and cognitive ability influence responses to biodata questions in a high-stakes selection context. *Personnel Psychology*, *65*(2), 359–383. https://doi.org/10.1111/j.1744-6570.2012.01245.x

Levine, T. R. (2014). Truth-Default Theory (TDT): A theory of human deception and deception detection. *Journal of Language and Social Psychology*, *33*(4), 378–392. https://doi.org/10.1177/0261927X14535916

Levine, T. R., Ali, M. V., Dean, M., Abdulla, R. A., & Garcia-Ruano, K. (2016). Toward a pan-cultural typology of deception motives. *Journal of Intercultural Communication Research*, *45*(1), 1–12. https://doi.org/10.1080/17475759.2015.1137079

Levine, T. R., Park, H. S., & McCornack, S. A. (1999). Accuracy in detecting truths and lies: Documenting the 'veracity effect.' *Communication Monographs*, *66*(2), 125–144. https://doi.org/10.1080/03637759909376468

Lewin, K. (1947). *Group decision and social change.* In T. Newcomb & E. Hartley (Eds.), *Readings in Social Psychology*, (pp.197–211). Holt, Rinehart & Winston.

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods*, *49*(2), 433–442. https://doi.org/10.3758/s13428-016-0727-z

Lyon, T. D., Malloy, L. C., Quas, J. A., & Talwar, V. A. (2008). Coaching, truth induction, and young maltreated children's false allegations and false denials. *Child Development*, *79*(4), 914–929. https://doi.org/10.1111/j.1467-8624.2008.01167.x

Mann, S., Vrij, A., Shaw, D. J., Leal, S., Ewens, S., Hillman, J., Granhag, P. A., & Fisher, R. P. (2013). Two heads are better than one? How to effectively use two interviewers to elicit cues to deception. *Legal and Criminological Psychology*, *18*(2), 324–340. https://doi.org/10.1111/j.2044-8333.2012.02055.x

Masip, J., Blandón-Gitlin, I., de la Riva, C., & Herrero, C. (2016). An empirical test of the decision to lie component of the Activation-Decision-Construction-Action Theory (ADCAT). *Acta Psychologica*, *169*, 45–55. https://doi.org/10.1016/j.actpsy.2016.05.004

Masip, J., Sporer, S. L., Garrido, E., & Herrero, C. (2005). The detection of deception with the reality monitoring approach: A review of the empirical evidence. *Psychology, Crime & Law*, *11*(1), 99–122. https://doi.org/10.1080/10683160410001726356

McCornack, S. A. (1992). Information manipulation theory. *Communication Monographs*, *59*(1), 1–16. https://doi.org/10.1080/03637759209376245

McCornack, S. A., Morrison, K., Paik, J. E., Wisner, A. M., & Zhu, X. (2014). Information Manipulation Theory 2: A propositional theory of deceptive discourse production. *Journal of Language and Social Psychology*, *33*(4), 348–377. https://doi.org/10.1177/0261927X14534656

McCornack, S. A., & Parks, M. R. (1986). Deception detection and relationship development: The other side of trust. *Annals of the International Communication Association*, *9*(1), 377–389. https://doi.org/10.1080/23808985.1986.11678616

Methasani, R, Gaspar, J. P., Barry, B. (2017). Feeling and deceiving: A review and theoretical model of emotions and deception in negotiation. *Negotiation and Conflict Management Research,10(*3), 58–178. https://doi.org/10.1111/ncmr.12095

Metts, S. (1989). An exploratory investigation of deception in close relationships. *Journal of Social and Personal Relationships*, *6*(2), 159–179. https://doi.org/10.1177/026540758900600202

Miller, R. L., Brickman, P., & Bolen, D. (1975). Attribution versus persuasion as a means for modifying behavior. *Journal of Personality and Social Psychology*, *31*(3), 430–441. https://doi.org/10.1037/h0076539

Mjelde, J. W., Schwarzlose, A. A. I., Dudensing, R. M., Jin, Y., Cherrington, L. K., & Chen, J. (2016). Subject pool effects among the general population and students: A choice experiment example. *Applied Economics Letters*, *23*(14), 1018–1021. https://doi.org/10.1080/13504851.2015.1128071

Moon, H. (2001a). Looking forward and looking back: Integrating completion and sunk-cost effects within an escalation-of-commitment progress decision. *The Journal of Applied Psychology*, *86*(1), 104–113. https://doi.org/10.1037/0021-9010.86.1.104

Moon, H. (2001b). The two faces of conscientiousness: Duty and achievement striving in escalation of commitment dilemmas. *The Journal of Applied Psychology*, *86*(3), 533–540.

Moriarty, T. (1975). Crime, commitment, and the responsive bystander: Two field experiments. *Journal of Personality and Social Psychology*, *31*(2), 370–376. https://doi.org/10.1037/h0076288

Nahari, G., & Vrij, A. (2014). Can I borrow your alibi? The applicability of the verifiability approach to the case of an alibi witness. *Journal of Applied Research in Memory and Cognition*, *3*(2), 89–94. https://doi.org/10.1016/j.jarmac.2014.04.005

Nahari, G., Vrij, A., & Fisher, R. P. (2012). Does the truth come out in the writing? Scan as a lie detection tool. *Law and Human Behavior*, *36*(1), 68–76. https://doi.org/10.1037/h0093965

Nahari, G., Vrij, A., & Fisher, R. P. (2014a). Exploiting liars' verbal strategies by examining the verifiability of details. *Legal and Criminological Psychology*, *19*(2), 227–239. https://doi.org/10.1111/j.2044-8333.2012.02069.x

Nahari, G., Vrij, A., & Fisher, R. P. (2014b). The Verifiability Approach: Countermeasures Facilitate its Ability to Discriminate Between Truths and Lies. *Applied Cognitive Psychology*, *28*(1), 122–128. https://doi.org/10.1002/acp.2974

Nasiry, J., & Popescu, I. (2011). Dynamic Pricing with Loss-Averse Consumers and Peak-End Anchoring. *Operations Research*, *59*(6), 1361–1368. https://doi.org/10.1287/opre.1110.0952

Olekalns, M., Horan, C. J., & Smith, P. L. (2014). Maybe it's right, maybe it's wrong: Structural and social determinants of deception in negotiation. *Journal of Business Ethics*, *122*(1), 89–102. https://doi.org/10.1007/s10551-013-1754-7

Olekalns, M., & Smith, P. L. (2007). Loose with the truth: Predicting deception in negotiation. *Journal of Business Ethics*, *76*(2), 225–238. https://doi.org/10.1007/s10551-006-9279-y

Ott, M., Cardi, C., & Hancock, J. T. (2013, July). Negative deceptive opinion spam [Paper presentation]. In Proceedings of NAACL-HLT, (USA) Atlanta, Georgia (pp.407-501). Association for Computational Linguistics

Palena, N., Caso, L., Vrij, A., & Nahari, G. (2021). The Verifiability Approach: A Meta-Analysis. *Journal of Applied Research in Memory and Cognition, 10*(1), 155–166. https://doi.org/10.1016/j.jarmac.2020.09.001

Parker, A. D., & Brown, J. (2000). Detection of deception: Statement Validity Analysis as a means of determining truthfulness or falsity of rape allegations. *Legal and Criminological Psychology*, *5*(2), 237–259. https://doi.org/10.1348/135532500168119

Patel, P. (2009). Experts expect résumé fraud to rise, *IEEE Spectrum*, *46*(6), 24–24. https://doi.org/10.1109/MSPEC.2009.4977602

Paulhus, D. L. (1998). *Paulhus deception scales PDS: The balanced inventory of desirable responding-7 manual*. Toronto, ON: Multi-Health Systems Inc

Pepitone, J. (2012). Yahoo confirms CEO is out after resume scandal. Retrieved from https://money.cnn.com/2012/05/13/technology/yahoo-ceo-out/index.htm

Peters, G. M. (1987). The Use of Lies in Negotiation. *Ohio State Law Journal*, *48*, 50.

Porfido, C. L., Cox, P. H., Adamo, S. H., & Mitroff, S. R. (2020). Recruiting from the shallow end of the pool: Differences in cognitive and compliance measures for subject pool

participants based on enrollment time across an academic term. *Visual Cognition*, *28*(1), 1–9. https://doi.org/10.1080/13506285.2019.1702602

Raskin, D. C., & Esplin, P. W. (1991). Statement validity assessment: Interview procedures and content analysis of children's statements of sexual abuse. *Behavioral Assessment*, *13*(3), 265–291.

Raskin, D. C., & Steller, M. (1989). Assessing Credibility of Allegations of Child Sexual Abuse: Polygraph Examinations and Statement Analysis. In H. Wegener, F. Lösel, & J. Haisch (Eds.), *Criminal Behavior and the Justice System: Psychological Perspectives* (pp. 290–302). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-86017-1_19

Raskin, D. C., & Yuille, J. C. (1989). Problems in Evaluating Interviews of Children in Sexual Abuse Cases. In S. J. Ceci, D. F. Ross, & M. P. Toglia (Eds.), *Perspectives on Children's Testimony* (pp. 184–207). Springer New York. https://doi.org/10.1007/978-1-4613-8832-6_9

Recruitment and Employment Confederation. (2017). *Perfect match. Making the right hire and the cost of getting it wrong.* Retrieved from https://www.rec.uk.com/our-view/research/recruitment-insights/perfect-match-making-right-hire-and-cost-getting-it-wrong

Ross, J., & Staw, B. M. (1993). Organizational escalation and exit: Lessons from the Shoreham Nuclear Power Plant. *The Academy of Management Journal*, *36*(4), 701–732. https://doi.org/10.2307/256756

Rosse, J. G., Stecher, M. D., Miller, J. L., & Levin, R. A. (1998). The impact of response distortion on preemployment personality testing and hiring decisions. *Journal of Applied Psychology, 83*, 634–644. https://doi.org/10.1037/0021-9010.83.4.634

Schmit, M. J., & Ryan, A. M. (1993). The Big Five in personnel selection: Factor structure in applicant and nonapplicant populations. *Journal of Applied Psychology, 78*(6), 966–974. https://doi.org/10.1037/0021-9010.78.6.966

Schmitt, N., & Kunce, C. (2002). The effects of required elaboration of answers to biodata questions. *Personnel Psychology, 55*(3), 569–587. https://doi.org/10.1111/j.1744-6570.2002.tb00121.x

Schmitt, N., Oswald, F. L., Kim, B. H., Gillespie, M. A., Ramsay, L. J., & Yoo, T, (2003). Impact of elaboration on social desirability and the validity of biodata measures. *Journal of Applied Psychology, 88*(6), 979–988. https://doi.org/10.1037/0021-9010.88.6.979

Scott, P. J., & Lizieri, C. (2012). Consumer house price judgements: New evidence of anchoring and arbitrary coherence. *Journal of Property Research*, *29*(1), 49–68. https://doi.org/10.1080/09599916.2011.638144

Shu, L. L., Mazar, N., Gino, F., Ariely, D., & Bazerman, M. H. (2012). Signing at the beginning makes ethics salient and decreases dishonest self-reports in comparison to signing at the end. *Proceedings of the National Academy of Sciences*, *109*(38), 15197–15200. https://doi.org/10.1073/pnas.1209746109

Sporer, S. L. (2004). Reality monitoring and detection of deception. In P. A. Granhag & L. A. E. Strömwall (Eds.), *The Detection of Deception in Forensic Contexts* (pp. 64–102). Cambridge University Press. https://doi.org/10.1017/CBO9780511490071.004

Sporer, S. L., & Schwandt, B. (2006). Paraverbal indicators of deception: A meta-analytic synthesis. *Applied Cognitive Psychology*, *20*(4), 421–446. https://doi.org/10.1002/acp.1190

Staw, B. M., & Ross, J. (1989). Understanding behavior in escalation situations. *Science (New York, N.Y.)*, *246*(4927), 216–220. https://doi.org/10.1126/science.246.4927.216

Stawiski, S., Tindale, R. S., & Dykema-Engblade, A. (2009). The effects of ethical climate on group and individual level deception in negotiation. *International Journal of Conflict Management*, *20*(3), 287–308. https://doi.org/10.1108/10444060910974894

Steller, M. (1989). Recent developments in statement analysis. In *Credibility assessment* (pp. 135–154). Kluwer Academic/Plenum Publishers. https://doi.org/10.1007/978-94-015-7856-1_8

Steller, M., & Boychuk, T. (1992). Children as witnesses in sexual abuse cases: Investigative interview and assessment techniques. In *Children as witnesses* (pp. 47–71). John Wiley & Sons.

Steller, M., & Köhnken, G. (1989). Criteria-Based Content Analysis. In *Psychological methods in criminal investigation and evidence* (pp. 217–245). Springer-Verlag.

Strenta, A., & DeJong, W. (1981). The effect of a prosocial label on helping behavior. *Social Psychology Quarterly*, *44*(2), 142–147. https://doi.org/10.2307/3033711

Strömwall, L. A., Hartwig, M., & Granhag, P. A. (2006). To act truthfully: Nonverbal behavior during a police interrogation. *Psychology, Crime and Law*, *12,* 207–219. https://doi.org/10.1080/10683160512331331328

Strudler, A. (1995). On the ethics of deception in negotiation. *Business Ethics Quarterly*, *5*(4), 805–822. https://doi.org/10.2307/3857416

Talwar, V., Arruda, C., & Yachison, S. (2015). The effects of punishment and appeals for honesty on children's truth-telling behavior. *Journal of Experimental Child Psychology*, *130*, 209–217. https://doi.org/10.1016/j.jecp.2014.09.011

Talwar, V., Gordon, H. M., & Lee, K. (2007). Lying in the elementary school years: Verbal deception and its relation to second-order belief understanding. *Developmental Psychology*, *43*(3), 804–810. https://doi.org/10.1037/0012-1649.43.3.804

Talwar, V., Lee, K., Bala, N., & Lindsay, R. C. L. (2002). Children's conceptual knowledge of lying and its relation to their actual behaviors: Implications for court competence examinations. *Law and Human Behavior*, *26*(4), 395–415. https://doi.org/10.1023/A:1016379104959

Talwar, V., Lee, K., Bala, N., & Lindsay, R. C. L. (2004). Children's Lie-Telling to Conceal a Parent's Transgression: Legal Implications. *Law and Human Behavior*, *28*(4), 411–435. https://doi.org/10.1023/B:LAHU.0000039333.51399.f6

Thijssen, S., Wildeboer, A., van IJzendoorn, M. H., Muetzel, R. L., Langeslag, S. J. E., Jaddoe, V. W. V., Verhulst, F. C., Tiemeier, H., Bakermans-Kranenburg, M. J., & White, T. (2017). The honest truth about deception: Demographic, cognitive, and neural correlates of child repeated deceptive behavior. *Journal of Experimental Child Psychology*, *162*, 225–241. https://doi.org/10.1016/j.jecp.2017.05.009

Vallacher, R. R., & Wegner, D. M. (1987). What do people think they're doing? Action identification and human behavior. *Psychological Review*, *94*(1), 3–15. https://doi.org/10.1037/0033-295X.94.1.3

Vallacher, R. R., & Wegner, D. M. (2012). Action identification theory. In P. A. M. Van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of Theories of Social Psychology: Volume 1* (pp. 327–348). Sage Publications Ltd

Vallano, J. P., Evans, J. R., Compo, N. S., & Kieckhaefer, J. M. (2015). Rapport-Building During Witness and Suspect Interviews: A Survey of Law Enforcement. *Applied Cognitive Psychology*, *29*(3), 369–380. https://doi.org/10.1002/acp.3115

Valley, K. L., Moag, J., & Bazerman, M. H. (1998). `A matter of trust': Effects of communication on the efficiency and distribution of outcomes. *Journal of Economic Behavior & Organization*, *34*(2), 211–238. https://doi.org/10.1016/S0167-2681(97)00054-1

van Kleef, G. A., De Dreu, C. K. W., & Manstead, A. S. R. (2004). The interpersonal effects of anger and happiness in negotiations. *Journal of Personality and Social Psychology*, *86*(1), 57–76. https://doi.org/10.1037/0022-3514.86.1.57

Vanderhallen, M., & Vervaeke, G. (2014). Between investigator and suspect: The role of the working alliance in investigative interviewing. In R. Bull (Ed.), *Investigative interviewing* (pp. 63–90). Springer New York. https://doi.org/10.1007/978-1-4614-9642-7_4

van der Wijst, P., & Krahmer, E. (2009). *Friendship, Deception and Punishment in Negotiations* (SSRN Scholarly Paper ID 1488625). Social Science Research Network. https://doi.org/10.2139/ssrn.1488625

Van Swol, L. M., Malhotra, D., & Braun, M. T. (2012). Deception and its detection: Effects of monetary incentives and personal relationship history. *Communication Research*, *39*(2), 217–238. https://doi.org/10.1177/0093650210396868

Van't Veer, A., Stel, M., & van Beest, I. (2014). Limited capacity to lie: Cognitive load interferes with being dishonest. *Judgment and Decision Making, 9(*3), 199–206.

Verschuere, B., Bogaard, G., & Meijer, E. (2020). Discriminating deceptive from truthful statements using the verifiability approach: A meta-analysis. *Applied Cognitive Psychology, 35,* 374–384. https://doi.org/10.1002/acp.3775

Vrij, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities*. Wiley, Chichester, England.

Vrij, A. (2019). Deception and truth detection when analyzing nonverbal and verbal cues, *Applied Cognitive Psychology*, *33*(2), 160–167. https://doi.org/10.1002/acp.3457

Vrij, A., Fisher, R., Mann, S., & Leal, S. (2006). Detecting deception by manipulating cognitive load. *Trends in Cognitive Sciences*, *10*(4), 141–142. https://doi.org/10.1016/j.tics.2006.02.003

Vrij, A., Fisher, R., Mann, S., & Leal, S. (2008). A cognitive load approach to lie detection. *Journal of Investigative Psychology and Offender Profiling*, *5*(1–2), 39–43. https://doi.org/10.1002/jip.82

Vrij, A., Granhag, P. A., & Porter, S. (2011). Pitfalls and opportunities in nonverbal and verbal lie detection: *Psychological Science in the Public Interest*. https://doi.org/10.1177/1529100610390861

Vrij, A., Leal, S., Granhag, P. A., Mann, S., Fisher, R. P., Hillman, J., & Sperry, K. (2009). Outsmarting the liars: The benefit of asking unanticipated questions. *Law and Human Behavior*, *33*(2), 159–166. https://doi.org/10.1007/s10979-008-9143-y

Vrij, A., Leal, S., Mann, S., & Fisher, R. (2012). Imposing cognitive load to elicit cues to deceit: Inducing the reverse order technique naturally. *Psychology, Crime & Law*, *18*(6), 579–594. https://doi.org/10.1080/1068316X.2010.515987

Vrij, A., Mann, S. A., Fisher, R. P., Leal, S., Milne, R., & Bull, R. (2008). Increasing cognitive load to facilitate lie detection: The benefit of recalling an event in reverse order. *Law and Human Behavior*, *32*(3), 253–265. https://doi.org/10.1007/s10979-007-9103-y

Vrij, A., Nahari, G., Isitt, R., & Leal, S. (2016). Using the verifiability lie detection approach in an insurance claim setting. *Journal of Investigative Psychology and Offender Profiling*, *13*(3), 183–197. https://doi.org/10.1002/jip.1458

Walczyk, J. J., Griffith, D. A., Yates, R., Visconte, S. R., Simoneaux, B., & Harris, L. L. (2012). LIE detection by inducing cognitive load: Eye movements and other cues to the false answers of "witnesses" to crimes. *Criminal Justice and Behavior*, *39*(7), 887–909. https://doi.org/10.1177/0093854812437014

Walczyk, J. J., Harris, L. L., Duck, T. K., & Mulay, D. (2014). A social-cognitive framework for understanding serious lies: Activation-decision-construction-action theory. *New Ideas in Psychology*, *34*, 22–36. https://doi.org/10.1016/j.newideapsych.2014.03.001

Walczyk, J. J., Igou, F. P., Dixon, A. P., & Tcholakian, T. (2013). Advancing lie detection by inducing cognitive load on liars: A review of relevant theories and techniques guided by lessons from polygraph-based approaches. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00014

Walczyk, J. J., Mahoney, K. T., Doverspike, D., & Griffith-Ross, D. A. (2009). Cognitive lie detection: Response time and consistency of answers as cues to deception. *Journal of Business and Psychology*, *24*(1), 33–49. https://doi.org/10.1007/s10869-009-9090-8

Walczyk, J. J., Schwartz, J. P., Clifton, R., Adams, B., Wei, M., & Zha, P. (2005). Lying person-to-person about life events: A cognitive framework for lie detection. *Personnel Psychology*, *58*(1), 141–170. https://doi.org/10.1111/j.1744-6570.2005.00484.x

Walsh, D., & Bull, R. (2012). Examining rapport in investigative interviews with suspects: Does its building and maintenance work? *Journal of Police and Criminal Psychology*, *27*(1), 73–84. https://doi.org/10.1007/s11896-011-9087-x

Yuille, J. C. (1988). The systematic assessment of children's testimony. *Canadian Psychology/Psychologie Canadienne*, *29*(3), 247–262. https://doi.org/10.1037/h0079769

Zhang, J.-D., Liu, L. A., & Liu, W. (2015). Trust and Deception in Negotiation: Culturally Divergent Effects. *Management and Organization Review*, *11*(1), 123–144. https://doi.org/10.1111/more.12028

Zuckerman, M., DePaulo, B. M., & Rosenthal, R. (1981). Verbal and nonverbal communication of deception. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 14, pp. 1–59). Academic Press. https://doi.org/10.1016/S0065-2601(08)60369-X

Appendices

Appendix A: Verbal and Nonverbal Elementary Deception Modes (EDM) Observed, Named, and Exemplified by the Author for 3 Years.

Note: some examples provided below have been modified and neutralized from our real-life observations, while others concern public personalities and have been replicated identically from public sources. In those cases, ground-truth is not always known but it was assumed that these examples were deceptive for the sake of simplicity and for pedagogic reasons.

#	Category	EDM name	Description	Example
			Verbal modes	
1	Deceptive Omission	Narrative Omission	Non-disclosure or mention of important information while recounting or answering a question about past activities.	A husband told his wife he would go back home early from work today. He ends up arriving more than an hour and a half late. When his wife asks him what happened, he answers: "I had a lot of work today". Even though he indeed had an intense day at work, he fails to mention he still found some time to have a drink with his colleagues after work before heading back home.
2		Strategic Information Omission	Non-disclosure or mention of important information while describing or answering a question about factual elements.	A man is selling his car to a buyer for $20,000 (the current market price for this kind of used car if well-maintained). During the selling process, he never mentioned to the buyer that important parts of the vehicle had to be replaced within 2 months and would cost around $5,000.

#	Category	EDM name	Description	Example
3		Likely Event Omission	Non-disclosure or mention of important information likely to occur in the future. This EDM has a higher Justification Power than the "Strategic Information Omission" EDM because people can later say: "I did not know about that back then!"	A man is selling his car to a buyer for $20,000 (the current market price for this kind of used car if well-maintained). During the selling process, he never mentions to the buyer that important parts of the vehicle would likely have to be replaced in the near future and should cost at least $5,000.
4		Hidden Agenda	Non-disclosure or mention of one's motivations and vested interests.	Example 1: Colleague A advises Colleague B against accepting a new permanent position she was offered within the company, listing many (real) disadvantages related to the position. Colleague A omits to mention though that she herself has applied for the very same position months ago, because she does not want to leave the city she is currently living in and that only that position would allow her not to have to relocate. Example 2: David is renting a small apartment and is angry at the landlord who just asked him to leave the apartment in the coming 3 months. To get revenge, he plans on burning the apartment while supposedly at work. Therefore, he premediates his plan and subscribe to an insurance 2 weeks before burning the apartment, in order to be at the same time reimbursed for his personal belongings he will lose in the staged fire.

#	Category	EDM name	Description	Example
5		Small Steps Strategy	Having a third party accept (or refuse) a basic premise in order to later have them accept the real target request.	Example 1: A seller in a supermarket offers a client a free tasting sample of some cake he is selling. [In reality, the seller thinks he will have a better chance of selling the cake to the client when making his offer after (rather than before) the client has tasted the cake.] Example 2: A landlord is selling his apartment. When asked about the selling price by a potential buyer, he first gives an unacceptably high price ($900,000), then "as a gesture" lowers it significantly to $750,000. [In reality, $750,000 has always been the targeted selling price by the landlord.]
6		Mental State Omission	Non-disclosure or mention of one's emotions, feelings, or thoughts.	A Psychiatrist asks their patient during depression therapy: "so how are you feeling today?" The patient answers: "Better." [In reality, while the patient is in fact feeling better than during the previous session, they omit to say that they still feel deeply depressed and has had suicidal thoughts lately.]
7		Selective Response	When several questions asked within a single sentence, choosing to answer only the question(s) for which truth can be disclosed while omitting answers to the sensitive question(s).	A 20-year-old youngster used marijuana at a Saturday night party with friends, but did not drink any alcohol. When driving back home around 2am, his car is pulled over by the police, who asks him: "have you used any drugs or drunk alcohol tonight?" The youngster quickly answers: "Officer, I haven't drunk a sip of alcohol!"

#	Category	EDM name	Description	Example
8		Suspicion Genesis Prevention	Setting the stage in advance for avoiding even raising suspicion later on. The goal is not to be able to later justify one's actions, but to "reverse engineer" what might raise suspicions in the future and act in the present to avoid such suspicions to arise.	A husband is planning to kill his wife on a Tuesday (his day off), then to carry her body to the dump in a bag. One year in advance, he thus took the habit of going every Tuesday to the landfill and dump a bag of human size volume (and to be clearly seen doing so), so that the day of the murder he would not raise any suspicion while carrying a big bag with his wife's dead body.
9		Intent Omission	Non-disclosure or mention of one's intents about forthcoming events.	At the customs border, a traveler is asked: "what are you planning to do here in NYC?" The traveler answers: "I don't know yet, tourism, I'll see!" [In reality the traveler intends to try to organize a secret business meeting in NYC, if he can contact the counterpart once in the city.]
10		Silence as a Response	Not answering comments as if silently agreeing to the counterpart's statements. Note that this EDM is different from refusing to answer a question.	A journalist says to his guest: "you are a renowned expert in Neuroscience, with a PhD from Cambridge, right?" [the guest remains silent]. "You were also a consultant for the National Committee on Brain diseases. Dr. Dupont is our guest tonight". [In reality, Mr. Dupont never had a PhD in Neurosciences.]
11		False Alternative	Pretending to a counterpart that they only have some alternatives available to them, when you know they actually have other ones.	A mother telling her child: "either you go to college, or you will become a cleaner!"

#	Category	EDM name	Description	Example
12		Encompassing Statement	Making a sentence that gives the impression of encompassing several previous statements or questions at once. This EDM is particularly convenient because one can later say "sorry, maybe I was not clear, I was only agreeing to points A and B, not the other ones".	Example 1: – Luc: "Following our last call, I have spoken to our CFO and CTO. They both confirmed what I told you, i.e., we would need to see and audit your software code before initiating any serious discussion about the acquisition. My CTO suggested he comes and audits your code on site after signing an NDA. Does that sound good to you? Should we schedule another call soon to further discuss this?" – Michel: "Yes it is ok. We can schedule a call on Tuesday 6th if that works for you." [In reality, Michel is still not ok with Luc's company auditing his (confidential) code, and was only agreeing to scheduling another call, as he will later say during the new call with Luc.] Example 2: "I am asking in this letter for a funding to attend the NanoPhysics international conference, which represents a unique occasion for me to meet in person with world-renowned Prof. Feng and Prof. Lino, with whom I already had a chance to exchange by email in the past and with whom I intend to build strategic partnerships." [In reality, the person writing the letter only exchanged emails with Prof. Lino, not Feng, but uses an encompassing statement to inflate the contacts they have already established.]

#	Category	EDM name	Description	Example
13		Hidden Identity	Using the passive voice or vague designations in sentences to hide the identity of the person responsible for an action.	Example 1: "It has been decided to fire him." Example 2: "The knife has cut his hand." Example 3: – Angelina: “I did not receive your mail. Why is that?” – Dorothy: “The person in charge of sending it has forgotten to do it.” [In reality, the person in question is Dorothy herself.]
14	Falsification	Narrative Fabrication	Pure fabrication of past activities that never occurred. This EDM is generally limited to the crucial part of the story and embedded within an otherwise truthful statement.	A youngster invents a story about how his bus had mechanical issues to explain to his professor why he is more than 20 minutes late. [In reality, he just woke up too late.]
15		False Response	Answering a close-ended question by providing a false answer (e.g., replacing a Yes with a No, providing a false figure, etc.).	Example 1: – Police officer: “Did you take the money from the envelope?” – Suspect: “No.” [In reality, the suspect had taken the money]. Example 2: – Buyer: “What is your own purchasing price?” – Seller: “€912 per ton.” [In reality, the purchasing price is €910 per ton.]

#	Category	EDM name	Description	Example
16		False Fact	Pure fabrication of a fact. This EDM is generally limited to the crucial part of a larger statement and embedded within an otherwise truthful statement.	On December 4th, 2012, the then French Minister Jérôme Cahuzac is accused by the journal "Mediapart" of having had an illegal bank account in Switzerland until 2010. On December 5th, Cahuzac reacts in a radio interview to these allegations by stating: "I do not have, I have never had, an account abroad, not now, not before." Four months later, he acknowledged having lied and having had a bank account in Switzerland until 2009.
17		False Opinion	Pure fabrication of an opinion.	Example 1: Alex, a life-long “beef lover”, has been dating a girl recently and is having dinner at her place for the first time. He discovers that she is a vegetarian and hates people who, in her words, "could do any harm to animals just to eat them, knowing that there are alternatives to animal proteins". Alex agrees: "I feel just like you about the animal cause." Example 2: A person consults a lawyer about their upcoming trial. The lawyer says: "I think that if you hire me, we will win the trial." [In reality, the lawyer believes that the person has no chance at all of winning the trial.]
18		False Intent	Fabrication about one's own intentions.	A person saying to one of their contacts: "thank you for your contact information; I will definitely email you so you have mine too." [In reality, the person has no intention at all to share their contact information and stay in touch.]

#	Category	EDM name	Description	Example
19		Bad Faith Offer	Offering (to do) something that you know perfectly well will not be feasible or accepted by the counterpart, as a way to later be able to say something along the lines of "I offered to do it, unfortunately it turned out to be impossible".	In August 2018, French rap singers Kaaris and Booba, two rivals, met at the Paris Orly Airport and engaged in a physical fight. Both were then convicted with 18 months suspended prison sentences and a €50,000 fine. After the trial, the two rappers continued to challenge each other through social networks, leading them to consider organizing an official ring fight and sending each other a contract to formalize it. Booba ends up rejecting the contract because it stipulates it will be organized in Tunisia, to which Kaaris replied on a social network video: "You asked for a contract with the hour, the day, the date, and a Greek sandwich; we sent you a contract with the hour, the day, the date, and a Greek sandwich. Now you're shitting yourself ... and everybody sees it. I would also have wanted it to happen in Europe, but you know perfectly well that with our track record in Orly, no Prefect would have ever given his authorization.... The promoter has contacted my lawyer, he read the contract, and Tunisia was the best compromise." [Kaaris is thus accusing his rival of deliberately having proposed to organize a fight which he knew would be impossible to do in the first place, at least in France.]

#	Category	EDM name	Description	Example
20		Trial Balloon	Proposing an idea involving a third party's collaboration or consent to see how they react to the idea. If the party's reactions are acceptable, the idea is executed, otherwise it is claimed to just have been an idea.	Thomas wants to go to Kevin's apartment on Friday to watch a game with all their friends, but his parents-in-law are coming to his home on Friday for dinner. Thomas tells his wife: "By the way I forgot to tell you, Kevin invited me to watch the game with everyone on Friday." Thomas' wife answers: "No way! Did you forget my parents were coming for dinner on Friday?" Thomas: "Yeah yeah I know, I was just saying that for your information."
21		Telling a Lie to Get at the Truth	Making a voluntary false statement as if it was a known fact, in order to have our counterpart consider a (usually confidential) fact as public knowledge and start disclosing sensitive information about it.	A father asks his child: "So tell me son, what are you going to offer to your Valentine this week?" The Son: "Dad, I am not going to offer anything because I don't have a girlfriend." [In reality, the father was trying to act as if it was known that his son had a girlfriend just to know if he actually had one.]
22		Feigning	Verbal simulation of a mental state. This EDM is often used with the Nonverbal EDM "Simulation".	Example 1: To avoid the maximum sentence, a criminal is feigning to answer the jury's question nonsensically in order to look crazy (and thus not fully responsible for their actions). Example 2: Joe's grandfather is starting to become a little deaf. To avoid embarrassment, he often feigns hearing the question he was asked by answering with generic sentences. For example, one day Joe asked his grandfather: “Grandpa, do you want some more wine or would you prefer some water now?” Grandfather replies: “Sure, sure.” Example 3: A person saying: "I feel very sad right now." [Whereas they feel absolutely neutral at the moment.]

#	Category	EDM name	Description	Example
23		Liability Transfer	Transferring one's own responsibility regarding a choice, act, or decision to a third-party.	Company A is seeking to acquire Company B and makes an offer to Company B's director. The director later calls Company A and says: "Thank you very much for your offer. To be honest, I was personally in favor of your offer, but my board of investors did not follow me on this, so I have to decline." [In reality, the director never presented the offer to his board and just was not interested in it.]
24		Calumny / Fake News	False information spread in order to damage a person's or an entity's reputation.	Surgeon A claims: "A lot of people say that Surgeon B never obtained his degree in Medicine." [In reality, nobody says that but Surgeon A, in order to harm his rival's reputation.]
25		Confidentiality Argument	Refusing to answer a question by (falsely) stating that the requested information is confidential and cannot be disclosed.	– Journalist: "Mr. President, could you tell us about the latest police developments in the investigation of potential illegal financing of your campaign? – President: "I am not allowed to discuss or comment on the work of the police and justice in an ongoing case."
26		False Memory Induction Attempt	Attempt to induce a false memory about a prior event that allegedly occurred in a counterpart's mind.	– Ben: "You could have told me that your friend was coming home for the weekend, instead of me having to accept this fait accompli!" – Tressia: "But I told three times she was coming for the weekend, you forgot, remember!" [In reality, Tressia never told Ben anything about a friend coming for the weekend.]

#	Category	EDM name	Description	Example
27		False Joke	Making a statement (that is false) as a joke. This EDM often relies on an element of surprise and the Nonverbal EDM "Simulation".	A young mother's baby is just starting to pronounce his first word: Mama. As the baby is just learning to speak, he says "Mama" with a strange intonation, stressing the last syllable: MA-MAAAAAAAA: – Baby: “MA-MAAAAAA” – Mother (making fun of the baby by imitating him): “MA-MAAAAA” – Baby's dad, laughing: “Why are you making fun of my son?” – Mother: “I'm not making fun of him ... I'm admiring him!”
28		Prank	Playing a trick or marking a joke or a surprise intended to be funny or pleasant to the deceiver and/or the target. This EDM often relies on a surprise element to elicit amusement.	– Robin: “Dude! Look our lottery ticket! We just won the first prize!” – Emile: “Oh my God let me see ...” [Emile quickly grabs the ticket with trembling hands.] – Robin: “Got ya!”

#	Category	EDM name	Description	Example
29		Playing Innocent	Pretending to be naive, not to understand (e.g., what is being reproached), not to having seen/heard something, not to be aware that something was expected, or pretending to have genuinely thought something to be true.	TV reality candidate Hillary is the ex-girlfriend of Vincent, who is now in a relationship with Virginie. There were many tensions between the candidates, but they have now decided to make peace. The day after that decision, when alone, Hillary says to Virginie: "You know, I am sorry things went so far between us. I really do not have any feeling anymore for Vincent. *I know this must be hard for you though, as I am Vincent's first love. But if Vincent ever loves you the way he loved me, then jackpot for you! He will never leave you."* Virginie gets angry at Hillary for insinuating Vincent does not love her as much as he loved Hillary. When Vincent complains to Hillary about what she said to Virginie, Hillary says: *"I didn't mean to be mean! I was just telling the truth! I didn't think that would hurt her!"* When interviewed alone later on, Hillary reveals: *"I was playing the innocent, I just wanted to hurt this bitch!"*
30		Misspeaking Argument	Recognizing having said something (often controversial), but justifying it by saying that you misspoke and did not mean what was said.	*July 2018* – After bipartisan criticism of his public undermining of US intelligence agencies during a press conference with Russian President Vladimir Putin in Helsinki, US President Donald Trump sought to bring closure after more than 24 hours of recrimination by saying he had simply misspoken when he said in Finland that he saw no reason to believe Russia had interfered in the 2016 US election. “The sentence should have been, ‘I don’t see any reason why I wouldn’t, or why it wouldn’t be Russia,’” instead of “why it would”, Trump said.

#	Category	EDM name	Description	Example
31		Failing Memory	Pretending not to remember or to have forgotten some information requested.	In October 2013, international French soccer player Patrice Evra relates an incident he had with former player and now TV consultant Bixente Lizarazu during an interview: – P. Evra: “Lizarazu? I do not know what he has against me.... He is the only player, for my first selection in the French team, who did not shake my hand.... And I remember that because Thierry Henry was with me and he said to him, ’Hey Liza, here is the succession’. And he looked at him and said, ’Why, who told you I was already retired?’, pissed off.” – B. Lizarazu (later asked on the TV set to react to the interview and if he was ever summoned together with Evra for a game with the French national team): “Listen, I do not remember that ... but very sincerely after seeing what I saw, I would have liked to answer what I said, supposedly. I would have really liked.”
32		Escape Planning	Mentioning in the present time a reason explaining our answer in case we get caught in the future.	A fraudulent policyholder is asked by his insurer (suspecting a fraud) if he has subscribed to other health insurances from other insurers to cover his recurring health care, without telling them (which is illegal). The policyholder denies having other undeclared health insurances, but adds: "given my chronic psychological conditions, I do not exclude making errors and ask for kind understanding". [In reality, the policyholder has subscribed to other undeclared insurances and remembers it perfectly well.]

#	Category	EDM name	Description	Example
33		Emotional Double Causality	Recognizing feeling an emotion, but lying about the cause of this emotion.	Sarah goes out for a drink with her good friend, Maya. Maya seems worried, anxious. Sarah asks her what is going on. Maya says: "it's work…I have some issues at work." [In reality, Maya has just received results of medical exams that are not good and has not totally processed the bad news yet to share it with Sarah.]

#	Category	EDM name	Description	Example
34		Motivational Double Causality	Recognizing having done/said something, but lying about the cause and our motivations for doing so.	Example 1: Juliet and Anna are both attracted to Raphaël. Juliet has disclosed this to Anna, but Anna never expressed her feelings about Raphaël to Juliet. At a party, Juliet sees Anna talking for a long time alone to Raphaël. Later, she asks, seemingly a little bothered, why she talking to him and maybe even flirting with him. "I was just talking to him to become friends and later be able to introduce you to him!" Anna answers. Example 2: Inès gets angry at her partner, Jimmy: "Oh you decided to go out with your friends yesterday, huh. Well you know what? I was not going to, but now I will accept my friend's invitation to their single-person party tonight!" [In reality, Inès always wanted to go to the party.] Example 3: A policyholder is suspected of fraud by their insurance company, which therefore asks them for documents before proceeding to the reimbursement they are asking for. The policyholder answers: "You are asking for documents I do not have. Because of my manic-depressive psychosis I have a most rudimentary administrative follow-up on those kinds of documents." [In reality, the policyholder does not have the documents required because they are trying to defraud their insurance company.]

#	Category	EDM name	Description	Example
35		Bogus Excuse	Explaining having done something (reprehensible) by providing an excuse that is not the true reason for the action. This EDM typically includes sentences like "sorry I did not know", "I had not heard you say that", "I forgot", or "It was just a joke".	Example 1: Pierre promised Robert he would quote Robert's company name in the TV interview he was about to do. Pierre does not really want to and actually ends up not quoting Robert's company. When later asked about it, he apologizes to Robert by saying "I forgot to quote you; I'm so sorry!" [The "I'm so sorry" part of the response being EDM "Feigning".] Example 2: Joey has thrown his Christmas tree in a waste container that he was aware was a private one. Unfortunately, CCTV cameras recorded his car license plate and allowed the container's owner to find and sue Joey. During his trial, Joey argues: "I did not know this was a private waste container; I thought this was a public one as it was accessible from the street!" Example 3: Bob complained to his friends that Julia, his girlfriend, was a "real pain in the ass". When Julia hears about it, she gets angry at Bob, who replies: "Come on don't get angry, I was just joking!" Example 4: Charles is filling out a life insurance subscription form. The form asks if he has ever had a disease for which his medical doctor asked him to do complementary medical exams (which he has). Charles answers "No" to this question. When later confronted by his insurer, he argues that the disease in question was so insignificant he did not even believe it to be relevant.

#	Category	EDM name	Description	Example
36		Situation Reversal	Attempt to reverse a situation to one's advantage while being aware that the arguments used are in contradiction with one's own secret beliefs.	Karen thinks she might have caught an STD while (secretly) cheating on her boyfriend Peter three weeks ago. In the meantime, Karen also had sex with Peter. She anxiously consults her gynecologist, who confirms an STD and prescribes medication for her and Peter. When Karen tells Peter they might both have to take pills according to her gynecologist because she might have an infection, Peter becomes suspicious and questions Karen's fidelity. She says: "Look, if it turns out I have an STD, I would have caught it from you and that would mean YOU would have cheated on me! So don't even get me started on this!"
37		Victimization	Attempting to act as a victim of a situation in order to attract sympathy from our counterpart or the public. This EDM is often used with the Nonverbal EDM "Simulation".	A Parisian French citizen gets arrested by the local police in Marseille (the second largest French city after Paris, and a football rival) for driving too fast. When later recounting the story to his friends, he is adamant he was arrested just because his license plate revealed he was from Paris, and complains about this injustice for 15 minutes. [In reality, the Parisian knows he was arrested just for driving too fast.]
38		Intimidation / Threat	Attempting to make a counterpart feel anxious and threatened, generally to obtain some behavior from them.	An aggressor is pointing a gun at a woman and says: "Give me your bag, or I will kill you!" [In reality, the aggressor has never harmed anybody and does not intend to start doing so. He just wants the money.]

#	Category	EDM name	Description	Example
39		Normality Claim	Pretending to find something (that we do not find normal) to be normal, in order to justify one's actions.	Example 1: Kevin is with friends preparing for a party. He tells them that he wants to look good tonight in order to be attractive to girls. Unfortunately, his mobile phone was accidentally calling his girlfriend while he was talking to his friends, and she heard everything he said. When later confronted by her, he talks and acts as if it is totally normal to want to be attractive, seemingly totally assuming his words. Example 2: John has just been hired as a salesperson in a printer company. As part of his new job, he has to visit one of the company's clients, located 1 hour away from his home, next Thursday at 8am. When discussing this visit at the printer company with his supervisor, he says: "Perfect I'll go and visit this company next week. Can you recommend any good hotels near the client?" [In reality, John knows that he is expected to go directly from his home to the client on the day of the visit, and not generate unnecessary costs by booking a hotel the night before. Because he thinks it will be more comfortable for him, he makes it appear as if it is natural to ask for the nearby hotels, therefore implicitly taking for granted he is authorized by his supervisor to book a hotel for this visit.]

#	Category	EDM name	Description	Example
40		Inception Attempt	Attempting to implicitly introduce a thought or belief in someone else's mind, so that they believe the thought in question emanates from themselves. This EDM includes the usage of connotated words to influence the perception of the counterpart.	Example 1: Victor is an entrepreneur selling cosmetic products. His business relies on finding as many agents as possible to sell Victor's products for commission. Therefore, Victor is always trying to project success and huge financial results when talking about his job, to attract new potential agents, so that people get to believe his business is booming (when really it is not even profitable) and they would become rich too if becoming an agent for him. For example, when talking with his friends about a car a friend of him is considering buying, Victor says: "Well, I personally would never buy car below $50,000. Not comfortable enough." Example 2: A dictator calling his opponents "terrorists" when they call themselves "resistance fighters". The usage of such positive/negative connotations is designed to influence the audience perception of the designated other party.
41		Collective Versions Reconciliation	A group of people collectively agreeing on a same alibi and version of an event in case of later interrogation.	A group of thieves has committed a robbery. They agree that if caught, they should all say that they were at Matthew's apartment during the time of the robbery.

#	Category	EDM name	Description	Example
42		Alibi Intent	Preparing in advance a reason for having done something, in order to have an alibi if later getting under suspicion.	A spouse reconnects with a former male colleague, allegedly regarding an old case they worked on together. The colleague answers by suggesting they discuss it in person sometime soon. [In reality, the spouse did not have to contact the former colleague about the case, but just wanted to have an excuse to reconnect with him. This whole maneuver was solely designed so that if the spouse's partner came across the messages someday, she could say that this exchange was purely professional (or better: the partner would think that independently when reading the content of the messages!)

#	Category	EDM name	Description	Example
43	Misinter-pretation	A Priori Play On Words	Using words with more than one meaning or interpretation to deceitfully answer a question, so that if confronted later we can say that we were telling the truth because we meant something other than what was understood by the interrogator.	Example 1: On January 26th, 1998, US President Bill Clinton concluded his White House press conference by denying the Monica Lewinsky affair with the following statement: "I did not have sexual relations with that woman, Miss Lewinsky". When later having to make a deposition in front of a judge, he was asked "Have you ever had sexual relations with Monica Lewinsky, as that term is defined in Deposition Exhibit 1?" Clinton answered: "I thought the definition included any activity by [me], where [I] was the actor and came in contact with those parts of the bodies." [In other words, Clinton denied that he had ever contacted Lewinsky's "genitalia, anus, groin, breast, inner thigh, or buttocks", and effectively claimed that the agreed-upon definition of "sexual relations" included giving oral sex but excluded receiving oral sex. During his initial statement on January 26th, 1998, Clinton was thus performing an "A Priori Play On Words" EDM.] Example 2: Tim attends his friend Joe's first art exhibition. After looking at a few paintings, Tim is categorical: he has never seen anything so ugly! When trying to discreetly leave the exhibition, he comes across Joe, asking him with a nervous look in his eyes: “So ... what do you think?" To avoid offending his good old friend, Tim answers: "Joe ... I really wasn't expecting that!" And Joe leaves taking this as a compliment from his friend.

#	Category	EDM name	Description	Example
44		A Posteriori Play On Words	Using different possible interpretations of our words from a former statement to pretend we meant something other than what was understood (e.g., saying things along the lines of: "you understood me wrong, sorry about that, what I meant was…").	Example 1: During a visit to Greece on September 8th, 2017, French President E. Macron declared: "I am of an absolute determination, and I will cede nothing, not to the slackers, not to the cynics, and not to the extremes." Later confronted with the controversy about the terms used in this statement, the President team quickly reacted by telling journalists that the word "slackers" was not referring to the French people, but to the political leaders "that have not conducted reforms for fifteen years". Example 2: – Gregory's mother says to her son: "Greg, you said you would clean up your room, and it's still a mess!" – Gregory: "Well, I never promised!"

#	Category	EDM name	Description	Example
45		Play On Ambiguities	Playing on situational ambiguities to deny knowledge on a given topic of interest.	Example 1: A woman lives with a drug dealer, who never explicitly discussed his activities with her. When later interrogated by the police, she denies any knowledge of him engaging in illegal activities. The police then say that given their standard of living, she must at least have had some idea about his activities. Example 2: On November 18th, 2015, 5 days after terrorist attacks that took place in Paris, Jawad Bendaoud is interviewed by journalists as the police raid an apartment that he declares owning. Before being himself picked up by the police in the middle of the interview, he declared: "I learned that (the police) were in my home.... I was not aware they were terrorists.... I was asked a favor, I did a favor." [In reality, it later turned out that Jawad Bendaoud was illegally renting this abandoned apartment for some time to migrants, dealers, and pimps. Right after the beginning of the police raid on the apartment, he exchanged text messages with a friend saying: "The guys they come from Belgium, they ask me on which side to do the prayer, they tell me we are tired, we want to sleep, we spent 3 days of sons of bitches, 150 euros for three days, why they have not been to the hotel?... Even I found the guys suspicious ..." Later interrogated by the police, he finally admitted: "I doubted, there was something unclear, but I will not take 20 years (of prison) for that.... I suspected it, but I wanted the money."]

#	Category	EDM name	Description	Example
46		False Irony	Telling the truth in such an exaggerated fashion (word content and/or tone of voice and/or face) that the counterpart thinks that you must be joking.	A husband asks his wife about the doubts he has been having for a long time regarding her fidelity: "Are you cheating on me? Tell me the truth. I can take it." His wife looks at him straight in the eyes, and says: "Well of course darling I am cheating on you! I even see my lover every Monday when you are travelling!" [In reality, the husband's wife does indeed see her lover every Monday.]
47	Distortion	Euphemism	Understating facts or feelings or opinions about other people or things.	A new colleague has just joined Louis's department at work. After a few weeks only, Louis has formed a definitive opinion about his new colleague: he finds him arrogant, condescending and simply put: he hates him. Yet, when asked about him by other people in the department, he invariably answers: "With him? We get along okay."
48		Exaggeration	Overstating facts or feelings or opinions about other people or things. This is the opposite process of the "Euphemism" EDM.	A person saying that their apartment is worth at least $700,000, when really they believe it to be worth $500,000 maximum.
49		False Modesty	Understating facts or one's feelings or opinions about oneself, in order to appear humble and/or likeable.	A surgeon has developed a new revolutionary medical device, and is invited to present it at an international talk. He starts by saying: "Thank you for this opportunity to present our innovation. I say ‘our’ because this was really a collective work and I want to thank all our staff for their involvement in this project. Their project." [In reality, the surgeon believes he alone developed this innovation and just wants to seem humble in front of his audience.]

#	Category	EDM name	Description	Example
50		Bragging	Overstating facts or feelings or opinions about oneself.	*September 2018* –"I don't believe there has been any administration in the history of this country that has done more in 2 years – and we're not even up to 2 years yet – than our administration," US President D. Trump said during a campaign speech in Las Vegas, while reading a list of those accomplishments.
51		False Flattery	Amplifying the alleged qualities of someone else, generally to gain their favor. Contrary to exaggerations which are self-centered, in flattery the object of amplification is someone else. This EDM is called "false" flattery (as opposed to "flattery"), because in false flattery the qualities put forward to gain someone's favor are deliberately amplified by the deceiver (i.e., they do not really believe their target has such qualities in such proportions).	Someone regularly telling their boss how brilliant they are in order to get a promotion.
52		Self-Serving Guilt-Tripping	Pretending to believe that our counterpart should feel guilty about (not) doing something, in order to influence them to behave in one's own interest.	A child does not want to kiss their mother in front of school where their friends could see them. She argues: "Next time you say that, just remember that I'll be gone some day and you'll regret rejecting my kisses!"

#	Category	EDM name	Description	Example
53		Mirage	Making someone believe in future perspectives in order to fool and manipulate them.	A software startup has developed a new innovative technology and is approached by a big firm for a partnership. The big firm's representative says to the startup's CEO: "If we make this deal together, we will open our network to your startup and you will thus be able to distribute your solution throughout our entire international distribution network. But before that, as an international firm we have to check a few prerequisites from you because we cannot afford to distribute a bad product and risk our reputation. So first of all we will need you to provide us with the source code of your solutions so we can audit it and make sure everything has been properly done." [In reality, the big firm is only making the startup believe in huge opportunities for them with the sole purpose of getting access to the precious highly confidential code.]
54		True In General	Choosing to answer a question by saying something that is true most of the time (but not always), in order to later, if confronted, be able to justify oneself by arguing we were talking "on average".	Simon is an ex-smoker who quit a year ago. He has never smoked a cigarette since, except last night after a dinner with old friends (all heavy smokers) he had not seen for several years. When later asked by his girlfriend if he had smoked last night, Simon answers: "You know I don't smoke." [Technically, since Simon has not smoked in one year, he is correct in saying he "does not smoke". On the contrary, saying "I did not smoke yesterday" would have been a falsification lie.]

#	Category	EDM name	Description	Example
55		Out of Context and Partial Citations	Quoting someone's words without providing the context of the words cited, usually in order to make the statement more controversial or caricatural. This EDM also includes voluntarily truncated citations changing the perception of a statement.	On March 10th, 2003, President Jacques Chirac justified his refusal of a French military intervention in Iraq during a televised speech: "My position is that, whatever the circumstances, France will vote No because it considers tonight that there is no need to wage a war to achieve the goal we set ourselves, that is, the disarmament of Iraq." Shortly after the statement, both French and foreign media quoted only "My position is that, whatever the circumstances, France will vote No." [This slight modification was not without consequences as the phrase was used to prove that France was a pacifist country which refuses war in all circumstances, not only in the specific case of Iraq.]

#	Category	EDM name	Description	Example
56		Misleading Approximation	Generalizing one's responses to a question to avoid providing some specific information and often in order to portray an enhanced image of oneself.	Example1: – Company A: “In which countries do you currently sell your products?” – Provider B: “Well we are currently selling here in the US, but also in Europe.” [In reality, Provider B is only selling their products in the US and Germany but does not want to reveal that they are only selling in one country abroad: Germany.] Example 2: – Recruiter: “How much do you earn annually?” – Job applicant: “Approximately $200,000.” [In reality, the job applicant earns $165,000 annually and intentionally makes a strong rounding up here.]
57		Misleading Facts Concatenation	Quoting several factual elements that are all individually true but that collectively do not honestly depict reality, in order to suggest a vision of the world.	Paul Lightman, an innocent person, is on trial for the murderer of Miss Danis: – Prosecutor: “is it true that you were in the neighborhood of Miss Danis the night of the murder?” – Paul Lightman: “Yes.” – Prosecutor: “Is it also true that 1 week before the murder you argued with Miss Danis?” – Paul Lightman: “Yes, but ...” – Prosecutor: “...And that you said to that she deserved the worst?”

#	Category	EDM name	Description	Example
58		Question Manipulation	Twisting the words used in the question of a counterpart, in order to be better able to answer their question.	When interviewed on December 4th, 2017 by TV journalist Jean-Jacques Bourdin, French government deputy Hervé Berville is asked: – Bourdin: “Is it true that the Christmas bonus will be renewed?” – Berville: “The Christmas prime? I do not have any Christmas prime as a deputy.” – Bourdin: “No not you, but the French citizens. No? The Christmas prime, you don't know what that is?" – Berville: “You may be giving me some information, but I do not have ANY Christmas prime!" – Bourdin: “Do you know what the ‘Christmas prime’ is? The Christmas prime is an exceptional grant to help the most vulnerable families during the end-of-year holidays.” [In reality, Berville was trying to avoid saying the prime for vulnerable families was not going to be renewed.]
59		Statement Manipulation	Twisting the words used in a counterpart's statement, usually in order to make them say something more controversial or caricatural. This EDM includes extending the meaning of words of the counterpart beyond their natural limits.	On January 11th, 2012, during his new year wishes, right-wing deputy Bernard Accoyer declared about the coming Presidential election in France: "If we miss this meeting of responsibility and courage, the economic and social consequences could be comparable to those caused by a war." The statement was then transformed by the press into: " 'If the left-wing passes, consequences comparable to a war', according to Bernard Accoyer."

#	Category	EDM name	Description	Example
60		Intensity Manipulation	Distorting the intensity or the strictness of an event.	A client writes to their supplier about a delayed delivery: "Please send me the order by Monday at the latest, this is very urgent, final deadline!" [In reality, the client had anticipated some possible delays from the supplier and could still meet their own deadline if they received their order on Tuesday instead of Monday.]
61		Protagonist Manipulation	Switching the protagonists of a story/event.	Paolo let his son drive his car to coach him before his driving license exam the following week, even though his son was thus not insured by Paolo's car insurance. Unfortunately, they had a small accident with the car, and the reparations have been assessed to cost about $3,000. To avoid paying for the reparations out of his own pocket, Paolo calls his insurance and declares that he was the one driving the car (the rest of the story about the accident being unchanged in Paolo's account).
62		Third-Party Intent Manipulation	Impugning (false) motives to someone else.	Kelly says that Fanny has done something specifically to hurt her. [In reality, Fanny did not intend to hurt anyone and Kelly is only saying that to other people to harm Fanny's good reputation out of jealousy.]

#	Category	EDM name	Description	Example
63		Location Manipulation	Changing the location of a story/event.	French citizen Lionel lives at the border of France and Germany, and regularly drives to Germany to make his weekly shopping. Unfortunately one day, Lionel has a small accident with his car in Germany, and knows that a fee is applied by his insurer when he has an accident abroad. To avoid paying this fee, Lionel calls his insurance and declares that his accident happened in France (the rest of the story about the accident being unchanged in Lionel's account).
64		Time Manipulation / Backdating	Changing the timing of events (that really happened) within a statement.	Jim is invited by Anna for dinner this Friday, but does not really want to go. He answers that he has to check his diary and will confirm later. In the meantime, he is invited to another dinner on Friday with very good friends he enjoys seeing. Jim ends up saying to Anna that unfortunately he won't be able to attend her dinner as he has another dinner already scheduled that night.
65		Likely Event Assertion	Affirming that an event is happening, when in fact it is at best very likely to happen.	A TV host is organizing a new show. He knows that convincing prestigious guests to participate on the show depends on the fact that other prestigious guests have already confirmed their participation. Therefore, when inviting a new star to participate to the show, he mentions that TV star "Charles Solar" has already agreed to participate. [In reality, the TV host's first exchanges with Charles Solar were in fact promising, but Charles Solar has not officially accepted to participate to the new show yet.]

#	Category	EDM name	Description	Example
66		Coded Language	Referring to some words by using other words (or nonverbal emblems) instead. This mode is a manipulation of the meaning of words.	Two criminals use coded language to discuss all their illegal activities to deceive policemen in case their phones are tapped (e.g., calling drugs "bananas", etc.)
67	Deceptive Evasion	Dodging / Political Cant	Giving the impression to having answered a question, but actually saying nothing specifically related to the question. This EDM is often used with the Nonverbal EDM "Saturation".	Right-wing politician Eric Woerth was interviewed on December 11th, 2017, the day after Laurent Wauquiez's election to the presidency of his party, the Republicans. – Journalist: “Is this [election] a clarification of the Republicans' line?” – E. Woerth: “I'll say a banality, but the Republicans are right-wing. We must stop apologizing for being both in opposition and being at the same time right-wing. We are right-wing, we are a modern right-wing, I think alive, open, but firm on these beliefs. But it goes from the center-right to the traditional right and Laurent [Wauquiez] to gather all without losing his identity. Without losing what he says, the strength of what he says, the percussive power he has and that is necessary, absolutely necessary to find at once ... that we are audible otherwise we are not audible and at the same time some credibility among the French”

#	Category	EDM name	Description	Example
68		Implied Response	Duping a counterpart by apparently providing an answer that implies the response to their question, without actually specifically answering the question.	Example 1: – Journalist: "Do you have a hidden bank account in Switzerland?" – Politician: "That's a joke." [In reality, the politician does have a hidden bank account in Switzerland. By saying "That's a joke", he wants the journalist to infer that because the question is "a joke", he therefore does not have a hidden bank account.] Example 2: – Police officer: "Did you kill your wife?" – Suspect: "I have never hurt anybody!" [In reality, the suspect did kill his wife, and wants the police officer to deduce from his answer that not hurting anybody includes not killing someone. Notice though that technically, the suspect could have killed his wife without actually making her suffer.]
69		Cognitive Response	Providing an answer deduced by a cognitive operation (e.g., a routine), not out of memory. This EDM allows the deceiver to later say: "I only answered by deduction, I did not say I remembered specifically."	– Police officer: "Where were you on Monday March 26th?" – Suspect: "I must have been to the gym, since I go to the gym every Monday."

#	Category	EDM name	Description	Example
70		Misleading Insinuation	Misleading a counterpart by suggesting things we do not believe. This EDM has the benefit of allowing the deceiver to later say, if caught: "this is not what I meant back then when saying that".	A left-wing political candidate is facing a scandal that has emerged right in the middle of the campaign. He goes on TV to address this scandal: – Journalist: "Who do you think is behind the leaks regarding this affair?" – Politician: "Well, just look at whom this benefits and you will have your answer." [The politician is suggesting that his political opponents are behind the leaks, whereas he knows perfectly well that the police started this investigation before the campaign and that they would probably reach their conclusions around the election – at a very bad time for him.]
71		Far-fetched Comparison	Making a far-fetched comparison (we do not believe in) between something done or said and something else, in order to justify what has been done or said.	Diego is a young dad. One day, his father (a former surgeon) is looking after his baby for the afternoon. When Diego comes back to pick up his child, his father tells him that he had the baby taste raw tomatoes. Diego is not happy because his pediatrician told him to only introduce cooked vegetables to the baby for the moment. Diego's father replies by saying: "look son, I am well situated to tell you that doctors sometimes say nonsense. I remember one time a colleague wanted me to reopen a patient after a very difficult surgery just to broaden the curettage zone, even though reoperating the patient could have killed him!"

#	Category	EDM name	Description	Example
72		Qualifying Statement	Making a statement (typically answering a close-ended question) by providing a non categorial response (typically, any other answer to a close-ended question than a Yes or No). This EDM allows deceivers to later be able to say: "I did not say I was sure / categorical about the topic".	Example 1: A company asked its bank for a loan to launch a new product development. When about to launch the project, the Company gives one last confirmation call to its bank, since it has not received any written confirmation yet and has to launch the project quickly to meet its deadlines: – Company: "Do you confirm you are granting us the loan so I can launch our project right away?" – Bank representative: "Go ahead, it should be good!" Example 2: George Tenet, former head of CIA, was interviewed in May 2007 about the CIA's interrogation program: – Journalist: "Anybody ever died in the interrogation program?" – Tenet: "No" – Journalist: "You're sure of that?" – Tenet: "Yeah, in this program that you and I are talking about, no."
73		Responsibility Dilution	Answering a question addressed to us by replying in the name of our broader group rather that in our own name, thus diluting our responsibility or knowledge within the group. This EDM is often used when the broader group in question truthfully does not have knowledge of the topic at stake, but the interrogated individual has.	The brother of a criminal is interrogated by the police: – Police: "Did you know where your brother was on Friday?" – Criminal's brother: "Our family did not know anything." [In reality, the criminal's brother was the only member of his family to know where his brother was on Friday. Therefore technically, his family as a group did not know where the criminal was.]

#	Category	EDM name	Description	Example
74		Indignation	Pretending to be indignant or outraged by a question to avoid answering it specifically. This EDM is often used with the Nonverbal EDM "Simulation".	– Employer: "Did you steal the money from the cash register?" – Employee: "How dare you accuse me? After all those years I've been working here! I can't believe that." [In reality, the employee did steal the money from the cash register.]
75		Ridicule / Derision	Turning to derision an (often embarrassing) question to be able not to answer it specifically.	*January 2017* – A day after unsubstantiated reports that Russian intelligence had a video allegedly showing him with prostitutes in 2013, in a hotel room where the women urinated on the bed, President Trump replied to those claims in the following manner: "Does anyone really believe that story? I'm also very much of a germaphobe by the way, believe me."
76		Irrelevant Proof	Providing irrelevant proof as an answer to a question, in an attempt to convince a counterpart.	A husband is suspected of infidelity by his wife. She asks him if he was with another woman last night. The spouse answers: "I was at the restaurant I swear; here is the bill you can have a look at it!"
77	Deceptive Diversion	Topic Switch	Changing the topic at stake in order not to answer a sensitive question. This EDM is convenient as it later allows, if caught, justification by saying: "I actually did not comment on that topic".	A journalist is interviewing a politician: – Journalist: "How do you feel about your indictment yesterday?" – Politician: "Look, I'll just say this. I find it very convenient that this indictment happens right in the middle of the campaign. How do you explain that? Do you really think there is no political involvement from the government here? Come on." [Notice how the politician changed the topic and never talked about his feelings to the indictment but diverted the discussion to accusations against the government.]

#	Category	EDM name	Description	Example
78		Diversion Joke	Using a joke not to address the topic at stake.	A husband is angry at his wife for not talking to him politely when she is under stressful circumstances: – Husband: "Are you like an animal, incapable of controlling yourself when under stress?" – Wife: "Yes I am an animal ... I'm your little kitty! Don't be mad."
79		Exasperation Card	Pretending to be bothered or exasperated by a question to be able not to answer it specifically. This EDM is often used with the Nonverbal EDM "Simulation".	A wife asks her husband: "Who are you having dinner with tonight?" The husband answers: "Oh come on stop bothering me with all your questions every time I have a dinner!"
80		Artificial Disturbance Creation	Trying to distract our counterpart from the (unfavorable to us) topic at stake by creating disturbances artificially.	A politician responding to a sensitive question during a debate by being disrespectful and trying to make his opponent angry in the hope that the debate switches to an argument about the politeness required in a gracious debate.
81		Strategic True Flattery	Answering a question by complimenting someone on qualities we really believe the counterpart has, in order not to answer a sensitive question.	– Juliet: "Are you lying to me right now?" – Romeo: "I never lie to beautiful girls."
82		Time Buying	Answering a question by using words and sentences only pronounced to buy time and think of an answer. This EDM is then followed by another EDM.	– Police officer: "Did you kill your wife?" – Suspect: "*Did I kill my wife?* How dare you. No." [In reality, the suspect did kill his wife. Notice how the suspect answered the question starting with the Time Buying EDM ("Did I kill my wife?"), then used the Indignation EDM ("How dare you") and ended with the Falsified response EDM ("No").]

#	Category	EDM name	Description	Example
83		Waters Muddying	Attempting to say something critical or controversial but at the same time making it go unnoticed, usually by diluting the point among many other noncontroversial ones listed, in order to later be able to tell the counterpart that we had already notified them about it.	A husband is listing to his wife all his forthcoming business travels, and quickly mentions *mezza voce* a personal trip with friends in the middle of the list, as he knows she would not be happy about him making this personal trip without her, given all the business travel his had to do already for his job.
84		Strategic Check	Answering a question by asking what exactly the person is talking about in order to evaluate their level of information and adapt the response accordingly, only commenting on the information disclosed by the counterpart.	Maria is married to Tom but having an affair with Denis. One day, Tom arrives at home furious: – Tom: "Admit it, I know everything!" – Maria: "And what is it you think you know?" – Tom: "You scratched the car and didn't tell me anything!" [In reality, Maria was afraid that Tom heard about her affair and wanted to check what his was talking about before responding.]
85		A Posteriori Cover Up	Following up on something one has just said (and regrets having said) by repeating the statement with slight differences in sonority as if correcting oneself from what was just a speech error.	A 20-year-old boy is talking to his girlfriend in a salacious manner in front of her parents' house: "I have very vicious intentions with you tonight (he realizes his girlfriend's father is in the garden looking at them) …so yeah, very serious intentions of working with you all night for the exam!"
86		Explanatory Diversion Creation	Feeling that certain unintentional behaviors might arouse suspicion, so trying to create a situation justifying those behaviors.	Alfred is playing a trick on his friend Zoe and pretending to have won $500 in the lottery. While doing so, he starts feeling slight smiles appearing on his face. In order not to raise Zoe's suspicions during his lottery story, he changes the topic and starts talking about a funny story to justify his increasingly broader smiles.

#	Category	EDM name	Description	Example
			Nonverbal modes	
87	Falsification	Simulation	Simulating an emotion or state that we do not feel in reality.	Example 1: Pretending to laugh to a joke we do not find funny at all (i.e., simulating amusement). Example 2: Playing dead to reduce the threat of an approaching bear.
88		Masking	Covering an emotion (often a negative one) with another one (often by a smile).	Masking one's sadness with a smile.
89		Disguise	Disguising oneself (by using make-up, costumes, etc.) to convey a falsified image of oneself.	A clown wearing a costume.
90		Forgery	Creating a false material proof.	A 16-year-old boy acquires a fake ID and shows it at the liquor store to be able to buy alcohol even though he is underage.
91	Misinter-pretation	Body Swerve	Trying to misdirect a counterpart with a body swerve.	A football player looking at the top right corner before shooting a penalty, in order to misdirect the goalkeeper and avoid him anticipating a shot close to the ground.
92		Implicit Induction	Duping a counterpart by apparently providing a nonverbal answer that implies the response to their question, without actually specifically answering the question (verbally or nonverbally).	A husband asks his wife: "Darling, do you still love me?" The spouse comes close to her husband and kisses him as a response. [In reality, she does not love him anymore.]
93	Distortion	Neutralization	Trying to suppress any nonverbal sign of emotions.	Adopting a "poker face" during a poker play.

#	Category	EDM name	Description	Example
94		Amplification	Exaggerating and amplifying a nonverbal attribute.	Example 1: Max was pranked by his friends who put a spider in his breakfast to scare him (Max hates spiders). When he first saw the spider in his bowl, he immediately jumped out of his chair in panic. While doing so, he realized what was going on, and in order for his friends not to make fun of him, he immediately starts to amplify his panic reaction, as if he was not afraid but pretending to be afraid. Example 2: While walking down a street at night, Julien comes across two scary people. To look strong and not be bothered by this group, he starts swaggering while walking in front of them.
95		Minimization	Minimizing and deamplifying a nonverbal attribute.	Example 1: Afifa and Emma are best friends and are both the finalists of an equitation contest. When the jury announced that Emma was the winner, she felt an intense happiness but managed her facial expressions to only show a slight smile, in order not to seem to be bragging in front of her good friend. Example 2: A thief is about to be caught by the police. Before being arrested, he discreetly throws the object he stole in a bush, with a minimal hand movement in order not to be detected.

#	Category	EDM name	Description	Example
96	Deceptive Diversion	Needle in the Haystack	Diluting a nonverbal behavior that we want to go unnoticed by artificially producing many other behaviors voluntarily. This EDM thus often involves the "Simulation" EDM to be produced after the behavior we want to go unnoticed to create the dilution.	A mother of two missing children is actually the killer of the children. Every time she feels she might have smiled in duping delight in front of the press, she starts crying ostensibly to make her smiles go unnoticed.
97		Concealment	Trying to conceal something sensitive (object, body part…) from the view of other people.	A terrorist hiding an explosive belt under a large coat.
98		Saturation	Creating a torrent of words (by adopting a high speech rate and/or loud voice volume) to divert the counterpart from the question they initially asked and avoid answering them specifically.	*See EDM "Dodging / Political Cant" and read the example imagining a high speech rate and loud voice from the respondent.*

Appendix B: General Theory of Deception's Deception Decision Algorithm Code

```
####    DEFINITIONS    ####

t_processing = the maximum available processing time of the solicited person remaining for the cognitive evaluation of the solicitation and best answer to provide
t_answer = the maximum time at which the solicited person thinks they have to answer the solicitation
t_end = min(t_processing, t_answer)
# the time at which the solicitation is made is considered to be 0

B_ND(i) = the evaluation of the benefit of not disclosing relevant, specific and expected item i with the soliciting person at t_answer. When there is an item i so that B_ND(i) > 0, it automatically launches the evaluation of the B(i), P(i), R(i), E(i), D(i) factors and their coefficients

B(i) = the evaluation of the benefit of answering with item i
P(i) = the evaluation of the punishment of answering with item i in case of an undesirable outcome
p(i) = the evaluation of the weight of P(i)
R(i) = the evaluation of the risk of being disbelieved when answering with item i
r(i) = the evaluation of the weight of R(i)
E(i) = the evaluation of the execution cost of answering with item i
e(i) = the evaluation of the weight of E(i)
D(i) = the evaluation of the dissonance of answering with item i
d(i) = the evaluation of the weight of D(i)

with: p(i) + r(i) + e(i) + d(i) = 1

# All BPRED factors evaluations include the evaluation/reevaluation over time of the coefficients p,r,e,d in the Utility function and the Utility function score update

prob(i)= the evaluated probability of undesirable event i

Events_full(solicitation) = a list with the full sequence of mental events triggered by the solicitation in question, with the events sorted in the list by chronological timing of occurrence. The timing of occurrences are indicated in parentheses for each event in this list
# Example 1: if for a given solicitation, there is one relevant specific expected item (called "Green1") and one relevant non-specific item (called "Yellow1") only that can be activated in Memory, then for example we could have:
Events_full(solicitation) = [Green1 activation (1.0), Yellow1 activation (1.5), B_ND(Green1) evaluation completion (2.1) with B_ND > 0, B(Yellow1) evaluation completion (2.2), D(Yellow1) evaluation completion (3.0), Undesirable1_of_Yellow1 + its probability prob(Undesirable1_of_Yellow1) evaluation completion (4.5), P(Yellow1) evaluation completion (5.0), R(Yellow1) evaluation completion (5.2), E(Yellow1) evaluation completion (6.0), Green1+Yellow1 fabrication activation (6.5), Justification1_of_Undesirable1_of_Yellow1 activation (7.1),...]
```

```
Events(t,solicitation) = the sublist of Events_full(solicitation) containing all the events triggered by the solicitation
and that have already occurred at time t
# In Example 1, Events(2.0,solicitation) = [Green1 activation (1.0), Yellow1 activation (1.5)]

Truth_full(solicitation) = the item in Memory corresponding to the exhaustive truthful answer to the solicitation. It is the
item containing all the relevant specific expected items to the solicitation listed in Events_full(solicitation)
# In Example 1, Truth_full(solicitation) = [Green1] because there was only one relevant specific expected item in
Events_full(solicitation).
# Example 2: if Events_full contained three relevant specific expected items to the solicitation in total, named "Green1",
"Green2", "Green3" and activated at times 1.0 / 2.0 / 3.0 respectively, then we would have Truth_full(solicitation) =
[Green1 + Green2 + Green 3]

Truth(t,solicitation) = the item in Memory corresponding to the part of the full truthful answer Truth_full(solicitation)
that is already activated in Memory at time t and thus accessible in the mind of the solicited person. This corresponds to
the full truth the person can access / remember at time t
# In Example 1: Truth(0.5,solicitation) = [], meaning that at that time the person solicited does honestly not yet remember
anything related to the question
# In Example 2: Truth(2.5,solicitation) = [Green1 + Green2]

Activated_expected(t,solicitation) = a list with all the relevant specific expected items triggered by the solicitation and
already activated in Memory at time t
# In Example 2: Activated_expected(2.5,solicitation) = [Green1, Green2]

Activated_nontruthful(t,solicitation) = a list with all the non-truthful options the solicited person can think at time t of
giving at t_answer. This list only starts to get filled once at least one relevant specific expected item i was activated
and B_ND(i) > 0. This list can be composed at time t of any item or combination of items activated in Memory at time t,
except Truth(t, solicitation). When an item enters this list, it automatically triggers the evaluation of its BPRED factors
and coefficients. The timings of activation of the non-truthful options are indicated in parentheses for each item in this
list

Undesirables_i(t) = the list of all the undesirable sequences of events that can happen at t > t_answer if answering with
non-truthful option i at t_answer the solicited person can think of at time t. The probability of each event as assessed at
time t is indicated in parentheses in this list, as well as the timing at which the undesirable event considered was
activated and its associated probability of occurrence evaluated. This list is created as soon as item i enters
Activated_nontruthful(t,solicitation)
# In Example 1, Undesirables_Yellow1(5.0) = [Undesirable1_of_Yellow1 (prob(Undesirable1_of_Yellow1);4.5)]

Justifications_Undesirablej_of_i(t) = the list of all the sequences of justifications the solicited person can think at time
t of giving at t > t_answer in case after answering with non-truthful option i at t_answer, undesirable sequence of events
Undesirablej_of_i occurs at time t > t_answer. Those justifications can either be new non-truthful answers or admitting to
```

```
the truth at time t > t_answer when the undesirable event occurs. This list is created as soon as item Undesirablej_of_i
enters Undesirables_i(t)
# In Example 1, Justifications_Undesirable1_of_Yellow1(7.1) = [Justification1_of_Undesirable1_of_Yellow1]

.select(t) = a function that searches in a list the items in their most advanced evaluation/scenario and selects the one
with the highest Utility score at time t
# Typically, in Example 1, the scenario involving deceptive answer Yellow1 is less advanced at t = 3.5 (Yellow1 activated &
only its B and D factors already evaluated) than at t = 7.1 (Yellow1 activated & all its BPRED factors evaluated & one
undesirable event than can occur if Yellow1 is answered at t_answer is considered (Undesirable1_of_Yellow1) and its
probability evaluated (prob(Undesirable1_of_Yellow1)) & one justification of it is already considered
(Justification1_of_Undesirable1_of_Yellow1 activation)
# NB: therefore the Utility function score of a given item can increase or decrease over time as each of its BPRED factors
progressively gets evaluated, or that more undesirable events come to mind to the solicited person, etc.

.score = a function that returns the value of the Utility function score of a given item in the Activated Memory


Answer(t) = the answer chosen at time t by the solicited person to be told at t_answer




####    UTILITY FUNCTIONS    ####

def UtilityFunction(option_t,t,solicitation):
    if option_t == Truth(t,solicitation):
        for i in Activated_expected(t,solicitation):
            U_i = B(i) - p(i)*P(i) - r(i)*R(i) - e(i)*E(i) - d(i)*D(i)
        return min(U_i,i in Activated_expected(t,solicitation))
    else:
        for i in Undesirables_option_t(t):
            for j in Justifications_Undesirablei_of_option_t(t):
                JustificationPower_j_i_option_t = B(j) - p(j)*P(j) - r(j)*R(j) - e(j)*E(j) - d(j)*D(j)
            JustificationPower_i_option_t = max(JustificationPower_j_i_option_t, j)
        u = 1 - sum(prob(i), i in Undesirables_option_t(t))
        return u*(B(option_t) - p(option_t)*P(option_t) - r(option_t)*R(option_t) - e(option_t)*E(option_t) -
d(option_t)*D(option_t)) + sum(prob(i)*JustificationPower_i_option_t, i in Undesirables_option_t(t))

# NB: the UtilityFunction is thus a prerequisite to the .select and .score functions that both implicitly rely on it
```

```
####   THE DDA ALGORITHM   ####


if t == t_end:
    for all items in Activated_expected(t_end,solicitation):
        if there is no item so that B_ND(item) > 0:
            The solicited person chooses at t_end to tell the truth at t_answer with truthful Answer(t_end) =
Truth(t_end,solicitation)

        else:
            if Events(t_end, solicitation).select(t_end).score > Truth(t_end,solicitation).score:
                    The solicited person chooses at t_end to answer non-truthfully at t_answer with nontruthful Answer(t_end)
= Events(t_end, solicitation).select(t_end)
            else:
                    The solicited person chooses at t_end to tell the truth at t_answer with truthful Answer(t_end) =
Truth(t_end,solicitation)
                    # Remark: in this case Truth(t_end,solicitation) = Events(t_end, solicitation).select(t_end)
```

Résumé français du contenu de la thèse

Théorie Générale du Mensonge :

Une théorie disruptive de la production, prévention et détection du mensonge

Présentation générale

La plupart des outils d'évaluation de la crédibilité décrits dans la littérature, comme le Criteria Based Content Analysis (CBCA ; Köhnken & Steller, 1988 ; Raskin & Esplin, 1991; Raskin et Steller, 1989; Raskin & Yuille, 1989; Steller, 1989; Steller et Boychuk, 1992 ; Steller & Köhnken, 1989 ; Yuille, 1988) ou le Reality Monitoring (RM ; Johnson & Raye, 1981 ; Masip et al., 2005 ; Sporer, 2004), requièrent des récits longs et libres des personnes interrogées pour distinguer menteurs et personnes sincères à des taux supérieurs à celui du hasard. Plus récemment, des chercheurs ont préconisé de passer d'une tentative d'observation passive d'indices comportementaux du mensonge (Buller et al., 1994, 1996 ; Ekman, 1985 ; Ekman & Friesen, 1969 ; Vrij et al., 2006), à l'adoption active de stratégies d'entretien conçues pour amplifier les différences de comportement entre menteurs et personnes sincères afin de mieux les discriminer. Par exemple, ce champ de recherche a montré que poser des questions exigeantes sur le plan cognitif aux personnes interrogées (par exemple, leur demander de raconter leurs activités passées dans l'ordre antéchronologique, leur poser des questions inattendues comme leur demander de décrire leurs activités sur un dessin, leur demander de répondre le plus rapidement possible aux questions, etc.) était jugé plus difficile à la fois par les menteurs et les personnes sincères, mais l'était beaucoup plus par les menteurs (Vrij et al., 2008, 2009, 2012 ; Walczyk et al., 2005, 2009, 2012). Cela a généralement conduit à rendre les différences de comportements (par exemple, la quantité de détails) plus saillantes entre menteurs et personnes sincères, améliorant ainsi leur taux de classification, par rapport à une tentative de classification lors d'un entretien passif.

Cependant, demander des récits ouverts et libres aux personnes interrogées, en particulier si ces demandes sont associées à des instructions d'entretien stratégiques, peut être inapplicable, inapproprié ou simplement jugé trop long dans des contextes d'entretien non judiciaires, qui sont pourtant les plus courants. Ainsi, ces stratégies ne s'appliquent pas aux besoins de professionnels cherchant à évaluer la véracité d'une déclaration factuelle (par opposition à l'évaluation de la crédibilité d'une déclaration portant sur la description d'activités passées). Dans un contexte d'entretien d'embauche par exemple, s'il est imaginable qu'un recruteur demande à un candidat de décrire ses tâches hebdomadaires typiques pour chacune de ses expériences professionnelles passées de manière aussi détaillée que possible (c'est-à-dire en demandant un récit ouvert), pour déterminer s'il travaillait réellement dans les entreprises en question, ce genre de stratégies d'entretien ne serait d'aucune utilité pour déterminer, par exemple, si le salaire annoncé par le candidat est véridique ou non.

Pour lutter contre de tels mensonges, courants dans la vie réelle mais peu étudiés dans la littérature, une nouvelle approche de l'évaluation de la crédibilité est proposée ici. Après la détection passive du mensonge, et l'adoption active de stratégies d'entretien pour une meilleure détection du mensonge, nous proposons maintenant de recourir en plus à des techniques qui tentent de dissuader les gens de mentir en premier lieu. Pour des mensonges portant sur des informations factuelles en particulier, ou portant sur de fausses dénégations, où aucune trace de mensonge peut n'être détectable, une approche alternative consiste à essayer de dissuader les gens de percevoir le mensonge comme leur meilleure option et de s'y engager. Mais qu'est-ce qui fait exactement penser aux gens que le mensonge est leur meilleure option, lorsqu'on les sollicite pour la vérité ? Quels facteurs psychologiques, ingrédients et « recette » appliquons-nous tous mentalement lorsqu'on nous demande la vérité afin de choisir de mentir ou non, et le cas échéant,

en utilisant quelle stratégie pour duper l'autre ? Si de tels facteurs existent et peuvent être identifiés de manière fiable, pourraient-ils être manipulés pour réduire la probabilité que des personnes agissent ou répondent de manière mensongère ? Et même si elles le font, la manipulation de ces facteurs pourrait-elle au moins rendre les indices du mensonge plus saillants, visibles et donc détectables ?

Dans le cadre du présent travail, une théorie générale du mensonge (General Theory of Deception, GTD) est d'abord proposée. Elle vise à unifier et à compléter les unités théoriques éparses qui ont été proposées dans la littérature sur le mensonge à ce jour, dans un modèle psychologique complet du mensonge, décrivant pleinement, de bout en bout, comment et quand les messages trompeurs sont produits, et comment cela permet de mieux prévenir et détecter le mensonge.

Une série d'expériences a ensuite été menée pour tester les prédictions de la théorie.

Collectivement, ce travail examine si la prévention et la détection du mensonge peuvent être améliorées en se fondant sur un cadre théorique complet.

La Théorie Générale du Mensonge (General Theory of Deception, GTD)

Comment mentons-nous ? Une observation sur trois ans de la production de mensonge dans des situations à faibles et forts enjeux

Nous définissons le mensonge et la tromperie (que nous utilisons indifféremment) comme « une tentative délibérée, réussie ou non, sans avertissement préalable, de créer chez un autre une croyance que le communicant considère comme fausse ». (Vrij, 2008, p. 15).

Pour construire la théorie GTD, nous avons commencé par identifier toutes les manières auxquelles un individu peut recourir pour tromper un tiers. Nous appelons toutes ces façons de mentir les « modes de mensonge » (Deception Modes, DM), qui sont une combinaison de

stratégies élémentaires de façons de mentir : les modes de mensonge élémentaires (Elementary Deception Modes, EDM). La liste des EDM est le résultat de l'observation, annotation et description quotidienne, pendant 3 ans, de la façon dont les gens produisent des messages trompeurs dans la vie réelle, verbalement ou non. Cette liste de 98 EDM a été divisée entre les EDM verbaux et non verbaux, et regroupée en six catégories : l'Omission (c'est-à-dire le fait ne pas divulguer des informations cruciales), la Falsification (énoncer quelque chose que l'on croit faux), la Mésinterprétation (compter sur la mauvaise interprétation de ses mots ou actions pour duper l'autre), la Distorsion (l'altération d'une base véridique), l'Evasion (ne pas répondre spécifiquement à une sollicitation ou question) et la Diversion (changer de sujet pour éviter de répondre). Quelques exemples d'EDM appartenant à chacune de ces catégories sont fournis dans le tableau ci-dessous.

Exemples de modes de mensonge élémentaires (EDM) parmi les 98 recensés pour chaque catégorie de mensonge.

Catégorie	Nombre total d'EDM	Exemple d'EDM	Description	Illustration
Omission	13	EDM 7 - Réponse sélective	Lorsque plusieurs questions sont posées au sein d'une seule et même phrase, choisir de répondre uniquement à la question pour laquelle la vérité peut être révélée, tout en omettant de répondre aux questions sensibles.	Un jeune homme de 20 ans a consommé de la marijuana lors d'une fête un samedi soir avec des amis, mais n'a pas bu d'alcool. En rentrant chez lui vers 2 heures du matin, sa voiture est arrêtée par la police, qui lui demande : « avez-vous consommé de la drogue ou bu de l'alcool ce soir ? ». Le jeune répond rapidement : « M. l'agent, je n'ai pas bu une goutte d'alcool ce soir ! »
Falsification	29	EDM 20 – Ballon d'essai	Proposer une idée impliquant la collaboration ou le consentement d'un tiers pour voir comment il réagit à l'idée. Si ses réactions sont acceptables, l'idée est exécutée, sinon elle est prétendue n'avoir été qu'une idée.	Thomas veut aller à l'appartement de Kevin vendredi pour regarder un match avec tous ses amis, mais ses beaux-parents viennent dîner chez lui ce soir-là. Thomas raconte à sa compagne : « Au fait, j'ai oublié de te le dire, Kevin m'a invité à regarder le match avec tout le monde vendredi ». Sa compagne lui répond : « Pas question ! As-tu oublié que mes parents venaient dîner vendredi ?" ». Thomas : « Ouais ouais je sais, je disais juste ça pour ton information ».

Catégorie	Nombre total d'EDM	Exemple d'EDM	Description	Illustration
Mésinterprétation	4	EDM 46 - Fausse ironie	Dire la vérité d'une manière tellement exagérée (contenu des mots et/ou ton de voix et/ou expression faciale) que la contrepartie pense que vous devez plaisanter.	Un mari interroge son épouse sur les doutes qu'il nourrit depuis longtemps quant à sa fidélité : « Est-ce que tu me trompes ? Dis-moi la vérité. Je pourrais tout encaisser ». Son épouse le regarde droit dans les yeux, et lui dit : « Bien sûr chéri, que je te trompe ! Je vois même mon amant tous les lundis quand tu pars en déplacement ! » [En réalité, l'épouse voit bien son amant tous les lundis]
Distorsion	20	EDM 47 - Euphémisme	Minimiser des faits, sentiments ou opinions concernant d'autres personnes ou choses.	Un nouveau collègue vient de rejoindre le service d'Alex au travail. Au bout de quelques semaines seulement, Alex s'est fait une opinion définitive sur son nouveau collègue : il le trouve arrogant, condescendant et tout simplement : il le déteste. Pourtant, lorsque d'autres personnes de son département l'interrogent à son sujet, il répond invariablement : « Avec lui ? On s'entend sans plus ».

Catégorie	Nombre total d'EDM	Exemple d'EDM	Description	Illustration
Evasion	10	EDM 68 – Réponse par implication	Duper son interlocuteur en fournissant une réponse qui implique apparemment la réponse à sa question, sans y répondre spécifiquement en réalité.	– Policier : « Avez-vous tué votre femme ? » – Suspect : « Je n'ai jamais blessé personne ! » [En réalité, le suspect a tué sa femme et veut que le policier déduise de sa réponse que le fait de n'avoir blessé personne inclut le fait de n'avoir tué personne. Notez cependant que techniquement, le suspect aurait pu tuer sa femme sans la faire souffrir]
Diversion	10	EDM 77 – Changement de sujet	Changer le sujet dont il est question pour ne pas répondre à une question sensible. Cet EDM est pratique car il permet plus tard, si la personne est confrontée, une justification du type : « En fait, je n'ai jamais fait de commentaire à ce sujet ».	Un journaliste interviewe un homme politique : – Journaliste : Que ressentez-vous suite à votre mise en examen d'hier ? – Politicien : « Écoutez, je dirai simplement ceci. Je trouve très pratique que tout cela se produise en plein milieu de la campagne. Comment expliquez-vous cela ? Pensez-vous vraiment qu'il n'y ait aucune implication politique du gouvernement ici ? Allons ». [Remarquez comment le politicien a changé de sujet et n'a jamais parlé de ses sentiments à l'égard de sa mise en examen, mais a détourné la discussion vers des accusations contre le gouvernement]

Combien de façons y a-t-il de mentir ? Dénombrement des modes de mensonge dérivés des EDM

Après avoir listé tous les EDM observés, la question de savoir combien de modes de mensonge existent, c'est-à-dire combien de combinaisons d'EDM peuvent être utilisées dans un épisode mensonger, devient pertinente. Au sein d'un épisode mensonger, on peut en effet n'utiliser qu'un seul EDM parmi les 98 disponibles, mais aussi deux, trois… ou 98 EDM parmi les 98 disponibles.

Mathématiquement parlant, la question de dénombrer les combinaisons de modes de mensonge possibles étant donné le (nombre minimum de) 98 EDM existants revient à résoudre et calculer la valeur de $\sum_{k=1}^{98}\binom{98}{k}$.

Cette somme est égale à $2^{98}-1$, qui vaut environ 10^{29} (soit $100 * 10^9 * 10^9 * 10^9$). En d'autres termes, cela revient à dire qu'il y a environ « cent milliards de milliards de milliards » de nuances de mensonge, ou de manières de mentir, possibles au sein d'un même épisode mensonger.

Quand les mensonges sont-ils produits ?

Le Modèle des Cinq Forces (Five Forces Model)

Maintenant que nous savons comment tous les mensonges sont produits, c'est-à-dire à travers une combinaison d'un ou plusieurs des 98 EDM, une question cruciale demeure. Quand décide-t-on de s'engager dans un mensonge ? Et quand on le fait, comment s'effectue la sélection du mode de mensonge jugé le meilleur ?

Contrairement aux propositions théoriques précédentes qui suggèrent principalement une simple analyse du « résultat attendu » du mensonge (Masip et al., 2016 ; Walczyk et al., 2014), nous affirmons que la décision de mentir dépend de cinq et seulement cinq facteurs en compétition (un processus que nous appelons le « Modèle des Cinq Forces ») : les bénéfices

perçus B, la punition perçue P, le risque perçu R, le coût d'exécution perçu E, et la dissonance perçue D. Lorsqu'une situation appelle à une réponse d'un individu, implicitement ou explicitement, on ne peut réagir que de l'une des trois manières suivantes : répondre honnêtement, de manière mensongère ou décider de ne pas répondre (soit en gardant le silence, soit en disant explicitement quelque chose comme : « je ne vais pas répondre à cette question »).

Par conséquent, dans la théorie GTD, nous introduisons des « fonctions d'utilité » de mentir U_L, dire la vérité U_T, et ne pas répondre U_{NA}, qui comparent en quelque sorte les coûts et les bénéfices de ces différentes options lorsque la vérité est sollicitée auprès d'un individu. Toutes ces fonctions d'utilité ont la forme de base suivante :

$$\begin{cases} U = B - p * P - r * R - e * E - d * D \\ avec : p, r, e, d \ les\ poids\ respectifs\ de\ P, R, E, D \\ \qquad et : p + r + e + d = 1 \end{cases}$$

Ce qui pourrait apparaître de prime abord comme une modélisation complexe des facteurs déterminant la décision de mentir ou de dire la vérité mérite quelques explications.

B_L représente d'abord les bénéfices perçus par un individu de recourir au mensonge, tandis que B_T représente les avantages perçus de dire la vérité. B_{NA} représente les avantages perçus de ne pas répondre. Les avantages du mensonge (également appelés « motifs de mensonge » dans la littérature) ont été examinés par plusieurs auteurs cherchant à les classer. Par exemple, l'analyse de réponses à des enquêtes ont montré que les avantages du mensonge peuvent être classés en fonction du bénéficiaire du mensonge, c'est-à-dire (a) les bénéfices pour soi, (b) les bénéfices pour les autres, (c) les bénéfices à préserver une relation (Hample, 1980; Metts, 1989).

Du côté des coûts des fonctions d'utilité, P_L, P_T, and P_{NA} représentent respectivement l'importance perçue par un individu des punitions/conséquences négatives en cas d'issue indésirable des options de mentir, de dire la vérité ou de ne pas répondre à une sollicitation.

Deuxièmement, R_L représente le risque perçu par un individu d'éveiller des soupçons ou de ne pas être cru lorsqu'il prononce son mensonge, tandis que R_T représente le risque perçu d'éveiller des soupçons ou de ne pas être cru malgré le fait de dire la vérité. R_{NA} représente le risque perçu d'éveiller des soupçons lorsqu'on ne répond pas à une sollicitation. L'évaluation subjective de ces risques d'éveiller des soupçons chez d'autres personnes nécessite la théorie de l'esprit, c'est-à-dire la capacité d'imaginer les pensées, sentiments, connaissances ou intentions des autres, mais aussi d'évaluer la probabilité que ses propres mensonges soient crus par une autre personne (DePaulo, 1992 ; Gombos, 2006 ; Talwar et al., 2007).

E_L, E_T, et E_{NA} représentent ensuite respectivement les « coûts d'exécution » de mentir, dire la vérité ou de de ne pas répondre à une sollicitation. Ces coûts d'exécution (DePaulo & Kirkendol, 1989 ; Ekman, 2009 ; Sporer & Schwandt, 2006 ; Vrij et al., 2011 ; Zuckerman et al., 1981) peuvent être liés à des coûts cognitifs (par exemple, la fabrication du mensonge lui-même), émotionnels (par exemple, se sentir coupable de mentir) ou de contrôle/inhibition (par exemple, inhiber des impulsions corporelles liées au fait de mentir, pour paraître honnête).

Enfin D_L, D_T, et D_{NA} représentent ce que nous appelons la « dissonance » perçue (en allusion à Festinger, 1957) d'un individu lorsqu'il dit un mensonge, admet la vérité ou ne répond pas à une demande de vérité. Dans ce contexte, la dissonance peut tout d'abord représenter le conflit mental auquel est confronté une personne lorsque certains de ses mots ou comportements sont incompatibles avec ses propres comportements, actions, déclarations, idées ou valeurs passés. Nous appelons ce premier type de dissonance la « dissonance interne ». Dans la théorie

GTD, la dissonance capture également les cas où les comportements d'un individu sont en contradiction avec les valeurs, la morale ou l'éthique de la société ou du groupe social dans lequel il vit. Bien que celles-ci aient pu ne pas être intériorisées comme ses propres valeurs personnelles, il peut les respecter. Nous appelons ce second type la « dissonance externe ». Parce que la dissonance externe n'est pas nécessairement intériorisée par un individu, nous prédisons que celle-ci conduit à des conflits mentaux moindres que la dissonance interne.

Tous ces facteurs « BPRED » (Bénéfice, Punition, Risque, Exécution, Dissonance) représentent la perception et l'évaluation subjectives d'un individu du présent et du futur, au moment de l'évaluation. Dans les fonctions d'utilité, ces cinq facteurs ne sont pas nécessairement d'égale importance, selon les individus, le contexte ou le moment de l'évaluation. Par conséquent, ils sont pondérés dans les fonctions d'utilité par des coefficients qui capturent ces variations entre les individus et les contextes, et pondèrent les facteurs BPRED en conséquence.

Bien que toutes les fonctions d'utilité U ont la même forme de base décrite précédemment, les fonctions d'utilité des options de réponse non véridiques diffèrent dans la théorie GTD de celle de la vérité en ce qu'elles prennent en compte l'évaluation subjective d'éventuels événements indésirables qui peuvent survenir plus tard, après que l'individu a répondu de manière non véridique, et toutes leurs justifications possibles. Nous appelons l'évaluation résultante de toutes les justifications possibles d'une option non véridique son « Pouvoir de Justification », J. En d'autres termes, la fonction d'utilité d'une option de réponse mensongère N est modélisée comme suit :

$$\begin{cases} U_L^N = u * \overline{U}_L^N + \sum_i prob(i) * J_L^{i\ de\ N} \\ avec: u = 1 - \sum_i prob(i) \\ et: \overline{U}_L^N = B_L^N - p_L^N * P_L^N - r_L^N * R_L^N - e_L^N * E_L^N - d_L^N * D_L^N \\ et: J_L^{i\ de\ N} = max_j\left(J_L^{j\ si\ i\ de\ N}\right) \\ avec: J_L^{j\ si\ i\ de\ N} = B_L^j - p_L^j * P_L^j - r_L^j * R_L^j - e_L^j * E_L^j - d_L^j * D_L^j \end{cases}$$

Avec $\overline{U}_L^N$ représentant l'évaluation BPRED de répondre avec une option mensongère N au moment où l'individu pense avoir besoin de répondre à la sollicitation de vérité ($t_{réponse}$), $prob(i)$ la probabilité qu'un événement indésirable i se produise après avoir répondu de manière mensongère avec N à $t_{réponse}$, et $J_L^{j\ si\ i\ de\ N}$ l'évaluation BPRED de justifier mentir avec N à $t_{réponse}$ avec un autre mensonge (ou séquence de mensonges) j à $t > t_{réponse}$ au cas où un événement indésirable i se produirait. Par conséquent, $J_L^{i\ de\ N}$ représente la meilleure (séquence de) justification(s) que l'individu puisse penser à donner plus tard au cas où les choses tourneraient mal (avec l'événement i) après son mensonge initial N, après avoir comparé toutes les justifications possibles j auxquelles il pourrait penser. Enfin, la fonction d'utilité globale U_L^N de mentir avec N, pondère toutes ces options par rapport à leur probabilité respective $prob(i)$.

L'algorithme de décision de mensonge (Deception Decision Algorithm, DDA)

Sur base de l'évaluation des facteurs BPRED de notre Modèle des Cinq Forces, les règles de décision que nous postulons sont utilisées par un individu chaque fois qu'il est sollicité pour la vérité afin de déterminer s'il va se livrer à un mensonge ou non, et si c'est le cas, avec quel mode de mensonge, sont développées. Nous appelons ce modèle de décision « l'algorithme de décision de mensonge » (Deception Decision Algorithm, DDA). Cet algorithme est détaillé dans l'annexe B.

Discussion

La principale prédiction et implication pratique de la théorie GTD est que la compréhension des mécanismes décrits par le DDA et le Modèle des Cinq Forces peut permettre à un intervieweur de manipuler un contexte et la formulation de ses questions pour rendre la personne interrogée plus honnête. Plus précisément, parce que la théorie GTD modélise qu'un individu fonde sa décision de mentir ou de dire la vérité en fonction de son évaluation subjective des facteurs BPRED de toutes les réponses potentielles auxquelles il peut penser en réponse à une sollicitation, une conséquence directe est que changer artificiellement la perception de l'individu des facteurs BPRED pourrait non seulement rendre les indices du mensonge plus saillants, mais pourrait même dissuader l'individu de s'engager dans un mensonge en premier lieu. Tandis que la littérature sur le mensonge s'est quasi-exclusivement concentrée à ce jour sur le fait d'augmenter la charge cognitive (facteur E dans la théorie GTD) pour exacerber les indices du mensonge (par exemple en posant des questions inattendues) ou pour dissuader les individus de s'engager dans un mensonge (Van 't Veer et al., 2014), la théorie GTD prédit plus largement que la probabilité qu'un individu choisisse de mentir peut être manipulée et diminuée par n'importe quelle combinaison des éléments suivants : augmenter les bénéfices perçus de dire la vérité, et/ou diminuer les bénéfices perçus de mentir, et/ou diminuer les facteurs PRED de dire la vérité, et/ou en augmenter les facteurs PRED du mentir.

Etude 1 : Le mensonge en négociation : rendre les gens plus honnêtes avec une approche à deux facteurs

Les négociations dématérialisées se multiplient et tendent à favoriser des échanges moins sincères, tout en limitant la capacité des négociateurs à détecter les signes de mensonge. Nous avons donc testé s'il était possible de dissuader les négociateurs de produire des déclarations

mensongères en manipulant le conflit mental du mensonge (facteur Dissonance) et en augmentant le risque perçu de ne pas être cru en cas de mensonge (facteur Risque).

Procédure

Un total de 458 participants a été recruté en ligne et placé dans un scénario de négociation acheteur/fournisseur réalisé par e-mail, avec huit groupes expérimentaux. Le facteur Dissonance a été manipulé chez les participants (en utilisant, ou non, l'étiquetage prosocial « vous avez la réputation d'être honnête et digne de confiance »), ainsi que le facteur Risque (trois niveaux d'ancrage de prix d'une prétendue offre alternative, et un groupe témoin sans ancrage). Tous les participants, en tant que fournisseurs, se sont finalement vus demander une information très sensible par l'acheteur : leur propre prix d'achat.

Résultats

Les résultats ont montré que la proportion de participants ayant révélé leur vrai prix d'achat était significativement plus élevée lorsqu'ils étaient exposés à nos facteurs de dissuasion du mensonge. Parmi les participants qui n'ont pas esquivé et ont répondu à la question, 38,1 % des participants de la condition contrôle (pas de dissonance, pas de facteur risque) ont révélé leur véritable prix d'achat, contre 80,8 % dans la condition combinant dissonance et niveau de risque le plus élevé.

Conclusion

Les professionnels peuvent ainsi utiliser des facteurs de dissuasion du mensonge simples à mettre en œuvre et fondés sur la théorie GTD, pour rendre leurs homologues plus honnêtes dans des négociations.

Etude 2 : Rendre les candidats à un emploi plus honnêtes avec un formulaire fondé sur la théorie : une étude de terrain

La plupart des employeurs déclarent trouver des mensonges dans les CV (Hayes, 2017), pourtant les employeurs qui reçoivent un grand nombre de curriculum vitae ne peuvent pas, de manière réaliste, effectuer des vérifications systématiques des références des candidats. Le but de cette étude était de développer un moyen rentable pour les employeurs à la fois de réduire, et de détecter plus facilement, les inexactitudes dans les dossiers des candidats, au stade de pré-entretien d'embauche.

Procédure

Nous avons collaboré avec une société de recrutement, ainsi qu'une société de vérification des références des candidats, pour tester un mécanisme permettant de lutter contre la fraude dans les CV. À l'aide d'un dossier de candidature personnalisé spécialement conçu pour à la fois *dissuader* et *améliorer la détection* du mensonge dans les candidatures, 27 candidats réels, postulant tous à des postes de commerciaux, ont été invités à remplir le formulaire personnalisé au stade du pré-entretien, après avoir soumis leur CV initial (en format libre).

Résultats

Au total, 269 éléments communs aux deux formulaires de candidature ont été vérifiés. Notre stratégie de dissuasion du mensonge a fonctionné puisque les inexactitudes ont été réduites de 23% sur les CV initiaux au format libre, à 11% dans les dossiers de candidature personnalisés ($p < 0{,}001$, $d = 0{,}9$). Le dossier de candidature personnalisé a également permis de détecter efficacement le mensonge, puisque pour les expériences professionnelles passées pour lesquelles les candidats n'ont pas fourni, bien que demandé, de coordonnées de points de contacts pour faciliter nos vérifications de références, seulement 39% ne contenaient aucune fausse information,

contre 77% qui étaient totalement exactes lorsque les coordonnées de points de contacts étaient fournies (p = 0,007, h = 0,78).

Conclusion

Cette nouvelle méthode fondée sur la théorie GTD pour rendre les candidats plus honnêtes et faciliter la détection des mensonges semble être un moyen efficace et rentable de lutter contre la fraude dans les CV.

Etude 3 : « Les menteurs fournissent moins de détails » … Et alors ? Comparaison d'instructions à deux rappels pour détecter le mensonge avec une méthode intra-sujet

La plupart des recherches sur le mensonge fournissent des résultats inter-sujets (par exemple, les menteurs font *en moyenne* des déclarations moins détaillées). Or ce genre de résultats peut s'avérer d'intérêt limité pour les professionnels cherchant à évaluer la crédibilité d'individus donnés, plutôt que de groupes. Cette étude examine les instructions optimales d'une stratégie de rappels multiples intra-sujet pour détecter le mensonge.

Procédure

Un total de 110 participants, réparti entre un groupe Mensonge et un groupe Vérité, a été placé au hasard dans quatre conditions d'entretien : deux instructions basiques demandant aux participants de rapporter tout ce dont ils se rappellent (1), un rappel de base suivi d'une instruction de focalisation périphérique (2), un rappel de base suivi d'une consigne de demande de détails vérifiables associée au protocole d'information (3), et deux rappels avec consigne de demande de détails vérifiables associée au protocole d'information (4). Tous les rappels ont été codés en termes de détails totaux et de détails vérifiables.

Résultats

L'interaction Groupe (mensonge, vérité) x Rappel (premier, second) n'était significative que dans la condition 3, les participants sincères fournissant plus de détails vérifiables dans le deuxième rappel que dans le premier. Une règle de décision simple intra-sujet en a été déduite, permettant un taux de classification des participants de 76,9%.

Conclusion

Les professionnels peuvent évaluer de manière optimale la crédibilité en utilisant deux rappels (rappel de base suivi d'une consigne de demande de détails vérifiables associée au protocole d'information) et en observant l'évolution des détails vérifiables entre les rappels.

Discussion générale

La Théorie Générale du Mensonge (GTD) est une nouvelle théorie disruptive de la production, prévention et détection du mensonge. Pour la première fois, une théorie unique et exhaustive du mensonge s'attache à décrire, de bout en bout, lorsque la vérité est sollicitée auprès d'un individu : (1) comment les modes de mensonge élémentaires (EDM) lui viennent à l'esprit et peuvent être combinés, (2) comment ils sont mentalement évalués les uns par rapport aux autres, et par rapport à l'option de répondre honnêtement (ou de refuser de répondre à la sollicitation), (3) comment les évaluations de ces options peuvent être modélisées comme s'appuyant sur 5 forces en compétition - les facteurs BPRED, (4) comment des limites intrinsèques ou contextuelles de temps et de capacité mentale peuvent conduire à des évaluations partielles de ces différentes options et à des décisions sous-optimales de l'individu, et (5) comment ce processus dynamique peut être décrit sous la forme d'un algorithme, qui peut ainsi être testé expérimentalement par rapport à ses prédictions.

Les évaluations des facteurs BPRED par un individu étant purement subjectives, une implication naturelle de la théorie GTD est qu'une tentative, par un intervieweur, de manipuler « artificiellement » ces facteurs pour essayer de changer leurs perceptions par un individu devrait conduire à un changement dans sa tendance à mentir. Plus précisément, augmenter les bénéfices perçus par un individu de dire la vérité, et/ou diminuer ses bénéfices perçus de mentir, et/ou diminuer ses facteurs PRED perçus de dire la vérité, et/ou augmenter ses facteurs PRED perçus de mentir devraient tous avoir pour tendance de dissuader les gens de s'engager dans un mensonge en premier lieu. Et même si l'individu le faisait quand même, nous prédisons que de telles manipulations devraient tout de même avoir tendance à rendre les indices du mensonge plus exacerbés et donc détectables, par rapport à l'absence de manipulation de ces facteurs. Pour confirmer ces prédictions qui découlent de la théorie GTD, nous avons mené une série de trois expériences manipulant différents facteurs BPRED et évaluant l'impact de ces manipulations sur la prévention et/ou l'amélioration de la détection du mensonge.

Dans notre première étude expérimentale, les participants ont pris part à une négociation en ligne menée par courrier électronique, agissant en tant que fournisseurs dans un scénario acheteur/vendeur. La manipulation de deux des facteurs de la théorie GTD (la Dissonance et le Risque de mentir) a été facilement opérationnalisée en ajoutant ou non une phrase dans le mail initial reçu par les participants. Par exemple, la manipulation du facteur Dissonance du mensonge a été effectuée en ajoutant simplement la phrase d'étiquetage prosocial « Je vous contacte aujourd'hui parce que vous avez la réputation d'être un fournisseur honnête et digne de confiance » à la fin de l'e-mail. Comme prédit par la théorie GTD, la manipulation des facteurs de Dissonance et de Risque de mentir a considérablement augmenté la proportion de participants choisissant de répondre honnêtement lorsqu'on leur a demandé leur propre prix d'achat en tant que fournisseur - une question pourtant très sensible.

Notre deuxième expérience, une étude de terrain examinant le mensonge chez des candidats réels à des offres d'emploi, a confirmé les effets de la manipulation des facteurs BPRED, cette fois pour prévenir et également faciliter la détection du mensonge. Dans cette étude les candidats, qui postulaient tous à des postes de commerciaux, ont initialement fourni leur CV (au format libre), qui a servi de référence du niveau d'inexactitudes avant toute manipulation des facteurs BPRED, au cabinet de recrutement qui a publié les offres d'emploi. Toujours lors de la phase pré-entretien, ils ont ensuite tous reçu un dossier de candidature personnalisé, que nous avons conçu pour manipuler les facteurs Dissonance (en demandant aux candidats au début du dossier de cocher et de signer une case Oui/Non certifiant qu'ils fourniront des informations entièrement exactes dans le document), Punition (en rappelant dans le dossier l'article de loi obligeant les candidats à fournir des informations de bonne foi) et Risque (en demandant aux candidats de fournir les coordonnées de points de contacts à leurs postes précédents pour une éventuelle vérification des références) liés au mensonge. En utilisant le formulaire de candidature personnalisé, les inexactitudes sont passées de 23% dans les CV initiaux au format libre, à 11 %. Le formulaire personnalisé a également permis de détecter efficacement le mensonge quant aux expériences professionnelles passées. En effet, lorsque les candidats ne fournissaient pas de coordonnées de points de contacts pour faciliter nos vérifications, alors que cela leur était demandé, seulement 39% des expériences professionnelles concernées ne contenaient aucune information inexacte, contre 77% lorsque les coordonnées de points de contacts étaient fournies, faisant ainsi de l'absence de la fourniture de points de contact un indice fiable de mensonge dans ce contexte.

Notre dernier test expérimental de la théorie GTD s'est concentré sur l'amélioration de la détection du mensonge, en manipulant un facteur plus couramment étudié dans la littérature : la charge cognitive. Dans la théorie GTD, la charge cognitive représente un sous-facteur du facteur

coût d'exécution (*E*), composé des coûts cognitifs, émotionnels et de contrôle/inhibition du mensonge. Dans cette expérience, les participants ont pris part à un scénario de type policier et ont choisi d'être sincères (en menant leurs activités normales) ou de mentir sur leurs activités passées (alors qu'ils commettaient en fait un simulacre de vol d'un bon d'achat de 10 euros). Tous les participants ont ensuite été informés qu'ils allaient être interrogés en tant que suspects d'un vol survenu alors qu'ils se trouvaient à proximité du bâtiment où le vol a eu lieu. Quatre jeux d'instructions différents ont été testés dans cette expérience pour manipuler la charge cognitive des participants devant raconter leurs activités passées. Parmi ceux-ci, un seul jeu d'instructions, demandant aux participants de raconter d'abord leurs activités avec une instruction basique de rappel libre (« racontez tout ce que vous avez fait »), suivie d'un deuxième rappel demandant aux participants d'inclure un maximum de détails vérifiables comme indicateurs de leur sincérité (consigne de demande détails vérifiables associée au protocole d'information), s'est avéré utile pour distinguer les menteurs des participants sincères. Cette expérience illustre donc que la manipulation des facteurs BPRED pour améliorer la détection du mensonge peut nécessiter des pré-tests et expériences pilotes de différentes variantes de manipulations pour produire efficacement le résultat escompté, avant un déploiement à grande échelle en vie réelle par des professionnels.

Bien que nous pensions que les présents travaux et l'élaboration de la théorie GTD soient une contribution importante et nécessaire à la littérature scientifique sur le mensonge (qui est un champ de recherche principalement athéorique), cette théorie disruptive ouvre également la voie à de nouvelles questions et à de futures recherches à mener. Les recherches futures dans le domaine devraient ainsi tout d'abord explorer comment décliner une méthode d'entretien systématique d'évaluation de la crédibilité à partir de la théorie GTD. Nous pensons que les

éléments suivants, à confirmer expérimentalement, représentent des pratiques d'entretien prometteuses découlant de la théorie GTD :

(10) *Définir le mobile*. La théorie GTD postule que lorsqu'une sollicitation explicite ou implicite de la vérité est faite à un individu, cela déclenche automatiquement chez l'individu une évaluation mentale initiale pour savoir si la non-divulgation des informations demandées à la partie sollicitante peut avoir un quelconque avantage pour l'individu interrogé. Si ce n'est pas le cas ($B_{ND} = 0$), alors la théorie GTD suppose que l'individu choisit de répondre honnêtement. En d'autres termes, l'analyse des mobiles potentiels des différentes parties prenantes fournit déjà des éléments contextuels d'appréciation de la crédibilité, indépendamment de toute analyse spécifique des déclarations des protagonistes.

(11) *Favoriser l'honnêteté. Faciliter la détection du mensonge*. Les personnes cherchant à évaluer la crédibilité d'un individu devraient commencer leur entretien en manipulant autant que possible les facteurs BPRED, pour essayer de rendre les gens plus honnêtes et les dissuader de s'engager dans un mensonge en premier lieu. A défaut, cela devrait au moins faciliter leur capacité à détecter le mensonge.

(12) *« Personne ne veut mentir ». Un nouveau principe de détection du mensonge*. Dans l'algorithme de décision de mensonge (DDA) de la théorie GTD, les différentes options qu'un individu compare mentalement lorsqu'une sollicitation de vérité lui est faite sont modélisées à l'aide de fonctions d'utilité. L'option avec le score le plus élevé est celle sélectionnée pour répondre à la sollicitation. Entre les options trompeuses mentalement considérées, nous postulons que les plus éloignées des modes de mensonge de type Falsification (parmi les catégories d'EDM définies dans la théorie GTD, à savoir l'Omission, la Falsification, la Mésinterprétation, la Distorsion,

l'Évasion et la Diversion), auront généralement des scores de fonctions d'utilité les plus élevés. GTD prédit que les Cinq Forces poussent les gens à opter pour la tromperie la plus proche de ce qu'ils croient être la vérité. Considérant que le codage des 98 EDM potentiels dans une déclaration, comme moyen de détection du mensonge, peut s'avérer peu pratique dans la vie réelle en raison du nombre trop élevé d'EDM à coder, nous proposons de réduire l'analyse d'une déclaration à un principe unique résultant de la théorie GTD. *Personne ne veut mentir*. Nous ne disons pas ici que personne ne ment. Plutôt que les gens, autant que possible, essaient lorsqu'ils ne veulent pas répondre totalement sincèrement à une question, de tromper leurs interlocuteurs tout en gardant l'impression consciente ou inconsciente qu'ils ne mentent pas vraiment.

(13) *Définir la suspicion*. Puisque je postule que *Personne ne veut mentir* et que les gens lorsqu'ils mentent ont tendance à minimiser l'écart entre leur réponse et la vérité ; le fait de définir, en tant qu'intervieweur essayant d'évaluer la crédibilité, la suspicion exacte de mensonge que l'on a au début de l'entretien est d'une importance primordiale. C'est aussi une condition préalable et nécessaire à l'application du principe *Personne ne veut mentir* comme moyen de détection du mensonge. A titre d'exemple simplifié, imaginez un policier demandant à un suspect : « Avez-vous tué votre épouse ? ». Ce dernier répond : « Je n'ai jamais fait de mal à personne ! ». En reconnaissant que les soupçons du policier sont de savoir si le suspect a tué ou non son épouse, et en appliquant le principe *Personne ne veut mentir*, celui-ci devrait immédiatement remarquer que le suspect pourrait avoir tué son épouse (la suspicion) tout en étant convaincu qu'il ne l'a pas fait souffrir en mettant fin à ses jours, ce qui

pourrait faire (inconsciemment ?) penser au suspect qu'il n'est pas en train de mentir avec une telle réponse, et réduire ainsi probablement son facteur Dissonance.

L'adaptation des étapes d'entretien aux types de mensonge à détecter représente également une question théorique cruciale. Bien que les éléments décrits ci-dessus puissent tous parfaitement être appliqués lors d'un entretien visant à détecter des mensonges concernant des activités passées, les autres types de mensonges peuvent ne pas pouvoir être appréhendés avec toutes les étapes suggérées, ou même nécessiter de nouvelles stratégies d'entretien. Plus précisément, nous considérons que le mensonge peut se diviser en six types, décrits ci-dessous :

(7) *Mensonges sur des activités passées* (par exemple, une personne mentant sur ses activités en réponse à la question « Que faisiez-vous samedi dernier entre 20h et 23h ?)

(8) *Mensonges sur des faits* (par exemple, une personne qui ment en réponse à « Quel est votre salaire actuel ? »)

(9) *Dénégations mensongères* (par exemple, une personne mentant en réponse à « Avez-vous, ou avez-vous à un quelconque moment eu, un compte bancaire en Suisse ? »)

(10) *Mensonges sur des activités futures* (par exemple, une personne mentant en réponse à « Qu'avez-vous l'intention de faire pendant votre séjour à New York ? »)

(11) *Mensonges sur des opinions* (par exemple, une personne mentant en réponse à « Que pensez-vous de ce groupe idéologique ? »)

(12) *Mensonges sur des motivations* (par exemple, une personne mentant en réponse à « Quelle est la raison de votre séjour dans le pays ? »)

Après une première phase historique dans la communauté scientifique de recherche de moyens passifs de détection du mensonge, suivie d'une phase de recherche de stratégies actives de détection cherchant à exacerber les indices du mensonge par augmentation de la charge

cognitive des sujets interrogés, la théorie GTD ouvre désormais une nouvelle voie de prévention même du mensonge. Plus généralement, de la recherche fondamentale aux applications réelles dans les entretiens conduits par des professionnels, la théorie GTD ouvre une myriade de nouvelles études passionnantes à mener au cours de la prochaine décennie.