



**HAL**  
open science

## Trust as indicator of robot functional and social acceptance. An experimental study on user conformation to iCub answers

Ilaria Gaudiello, Elisabetta Zibetti, Sébastien Lefort, Mohamed Chetouani,  
Serena Ivaldi

### ► To cite this version:

Ilaria Gaudiello, Elisabetta Zibetti, Sébastien Lefort, Mohamed Chetouani, Serena Ivaldi. Trust as indicator of robot functional and social acceptance. An experimental study on user conformation to iCub answers. *Computers in Human Behavior*, 2016, 10.1016/j.chb.2016.03.057 . hal-01298502

**HAL Id: hal-01298502**

**<https://inria.hal.science/hal-01298502>**

Submitted on 6 Apr 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# **Trust as indicator of robot functional and social acceptance.**

## **An experimental study on user conformation to iCub answers**

**Ilaria Gaudiello<sup>1</sup>, Elisabetta Zibetti<sup>1</sup>, Sébastien Lefort<sup>2</sup>, Mohamed Chetouani<sup>3</sup>, Serena Ivaldi<sup>4</sup>**

<sup>1</sup> CHART EA 4004, Univ. Paris 8 and EPHE Paris, LUTIN Living Lab – FED 4246, F-93526 Saint Denis, France

<sup>2</sup> LIP6, Sorbonne Univ. UPMC Paris 6, UMR 7606, F-75005 Paris, France

<sup>3</sup> ISIR, CNRS & Sorbonne Univ., UPMC Paris 6, UMR 7222, F-75005 Paris, France

<sup>4</sup> INRIA, Villers-lès-Nancy, F-54600 France & LORIA, CNRS & Univ. Lorraine, UMR 7503, Vandoeuvre-lès-Nancy, F-54500 France.

### **Abstract**

To investigate the dynamics of human-robot acceptance, we carried out an experimental study with 56 adult participants and the iCub robot. Trust in the robot has been considered as a main indicator of acceptance and measured by the participants' conformation to the iCub's answers to questions on functional and social tasks characterized by perceptual and socio cognitive uncertainty. In particular, we were interested in understanding whether (i) trust in functional savvy is a prerequisite for trust in social savvy, and (ii) to what extent factors such as participants' desire for control, attitude towards social influence of robots, and imagined collaborative *vs.* competitive scenario, may influence their trust in the iCub. We found that participants conformed more to the iCub's answers in the functional than in the social tasks. Moreover, the few participants conforming to the iCub's answers in the social task also conformed less in the functional issues. Trust in the robot's functional savvy does not thus seem to be a pre-requisite for trust in its social savvy. Finally, the examined factors did not influence the trust in iCub. Results are discussed with relation to methodology of human-robot interaction (HRI) research.

**Keywords:** iCub humanoid robot; functional acceptance, social acceptance, trust, conformation

## Introduction

A major barrier to the introduction of robots in public and private places is the end-user acceptance. In public places, applications such as robot-guides for museums, robotic actors or entertainment robots for commercial purposes have been successful mostly because of the novelty effect and the astonishment they evoked in the public. However, a more thorough assessment of users' stance towards an enduring presence of robots in private places to help with daily routines is still missing in the literature.

Even though several models of technology acceptance have already proved their validity in the last decade (Ben Messaoud, Kharrazi & Mac Dorman, 2011; Venkatesh, Morris, Davis *et al.*, 2003; Dillon, 2001) we cannot apply these models to robots straightforwardly. In fact, as an autonomous agent endowed with an embodied intelligence and decision-making capabilities, the robot might be seen as having more intrusive consequences in our life than other technologies (see for instance Salem, Lakatos, Amirabdollahian *et al.*, 2015a,b - where a domestic robot requests a person to throw the correspondence of another person into the bin). Consequently, acceptance is today a sensitive topic of investigation in HRI research (*e.g.*, Fridin & Belokopytov, 2014; de Graaf & Allouch, 2013; Young, Hawkins, Sharlin *et al.*, 2009; Heerink, Krose, Wielinga *et al.*, 2009; 2010; Kaplan, 2004; Riek & Robinson, 2008; Salvini, Laschi & Dario, 2010; Welch, Lahiri, Warren *et al.*, 2010; Weiss, Bernhaupt, Lankes *et al.*, 2009).

Through an attentive survey of the literature, we can identify six connotations of robot acceptance: representational, physical, behavioral, functional, social, and cultural<sup>1</sup> (see Gaudiello, 2015 – Phd dissertation). Among these dimensions, functional and social acceptance appear to play a key-role for effective human-robot interaction (Fridin *et al.*, 2014; Weiss, Bernhaupt, Lankes *et al.*, 2008; Zaad & Allouch, 2008; Picard & Daily, 2005).

Functional acceptance refers to the level of perceived ease of use, usefulness (Heerink *et al.*,

---

<sup>1</sup> *Representational acceptance* describes the influence of the mental models (stereotypes from science fiction, representation borrowed from preexistent mental models of technology, animals or children, etc.) on robot acceptance in terms of place attributed to the robot within common sense ontology and functions or roles assigned to a robot. *Physical acceptance* defines the embodiment features of the robot - its morphology (anthropomorphic, zoomorphic, machine-like), size (human or animal size vs. household appliance or toy size, etc.) materials (*e.g.*, organic vs. mechanic), etc. – that could impact robot acceptance in terms of likeability and credibility of the robot. *Behavioral acceptance* concerns those behaviors – for example proxemics (*i.e.*, intimate vs. public distance, direction of approach, etc.), and communication (*e.g.*, verbal, gestural, etc.) – that can affect acceptance in terms of believability and fluency of interaction. Finally, *cultural acceptance* refers to fundamental issues of a given culture, for example appetite for technology, social care, educational values, etc. which might alter acceptance in terms of intention to use and duration of use.

2009; Weiss *et al.*, 2009), accuracy (Schaefer, 2013; Beer, Prakash & Mitzner, 2011), and innovativeness (Rogers, 1995; Kaplan, 2005) of the robot. Social acceptance covers a variety of issues such as social presence (Riether, 2013; Hamill & Harper, 2006), perceived sociability – often overlapping with social abilities and social intelligence (Heerink, 2010), and social influence (Nomura, Kanda & Suzuki, 2006) of robots.

Interestingly, while representational (Gaudiello, Lefort & Zibetti, 2015) and cultural (Li, Rau & Li, 2010) acceptance continuously vary with relation to users' progressive familiarization to robots (Bartneck, Suzuki, Kanda *et al.*, 2007), and while we are likely to accept that a robot might have a rough physical aspect (Turkle, 2011), or to excuse the robot for an inappropriate behavior like we would do with a child (Young *et al.*, 2009), we are much less likely to accept a useless robot (Kaplan, 2005) and we can be disappointed by a robot that lacks social abilities (Heerink, 2010; Shaw-Garlock, 2009).

Today, robots with functional and social abilities can still be distinguished in two categories on the market (Shaw-Garlock, 2009), with distinct purposes and labels - *e.g.*, 'utilitarian' and 'hedonistic' or 'entertainment' robots (Lee, Shin & Sundar, 2011). But robotic industry is definitely moving towards robots which are both functional and social (Shaw-Garlock, 2009) and researchers have argued that users behavior will evolve accordingly: the more efficient a robot is, the more people would be likely to rely on them on all levels, including social level (Young *et al.*, 2009).

Moreover, it has been hypothesized that with increasingly realistic and socially capable robots the propensity for people to treat technologies in the same way they treat other people (Reeves & Nass, 1996) could become stronger, with hardly imaginable consequences on an ethic level (Salem *et al.*, 2015a). Therefore, in order to better understand how robots' acceptance is psychologically grounded and to examine the related advantages and risks, contemporary studies have narrowed the research down to a fundamental common ingredient of functional and social interaction with robots: *users' trust in the robot* (Cameron, Aitken, Collins *et al.*, 2015; Salem *et al.*, 2015a, 2015b; Hancock, Billings, Schaefer *et al.*, 2012; Shinozawa, Naya, Yamato *et al.*, 2005).

Trust is increasingly employed to assess the quality of human-robot interaction (van den Brule *et al.*, 2014; Schaefer, 2013; Kaniarasu *et al.*, 2012; Yagoda, 2011; Hoffman, Lee, Woods *et al.*, 2009; Desai, Stubbs, Steinfeld *et al.*, 2009 Lee & See, 2004) and, as it emerges from this literature, is considered a valid indicator of robot's functional and social acceptance (Salem *et*

al., 2015a-b; Yagoda, 2011). In particular, recently Salem et al., (2015a) state that « Especially with regard to critical decisions, trust plays an important role in human interactions and could therefore help to increase the robot's acceptance in its role as a collaborative partner (Lee, Knox, Wormwood et al., 2013; Hinds, Roberts & Jones, 2004). Since trust is strongly linked to persuasiveness in social interaction contexts (Touré-Tillery & McGill, 2015), it could also affect people's willingness to cooperate with the robot (Freedy, de Visser, Weltman et al., 2007), for example, by accepting information or following its suggestions » (p. 585).

For example, suppose a person is asking its domestic robot to help in deciding which basket of tomatoes is the heaviest. In this case the person would have to trust the robot as a functional agent (*functional acceptance*). Now, suppose the same person is asking to the robot which is the indispensable equipment to bring in a swimming pool-bag of his child (a bathing-cap? A pair of flip-flops?). In this case the person would have to trust in the robot social answer (*social acceptance*). The question is: will the person trust the robot's answer in both of these cases? As showed in our example, functional and social acceptances are characterized by the same dynamics, but apply to very different issues. Since they are grounded in different abilities (technical estimation vs. social skills for human care), they presuppose very different built-in knowledge and robot performance.

However, the research models developed in the last decade to assess functional acceptance of robots (Fridin et al., 2014; Beer et al., 2011; Heerink et al., 2009; Weiss et al., 2009; Beer et al., 2011) mainly borrow their conceptual frame from older technology acceptance models (Sun & Zhang, 2006; Venkatesh, Morris, Davi et al., 2003; Rogers, 1995; Davis, Bagozzi & Warshaw, 1989) thus neglecting the specificity of the robot as a social agent. Several studies validated a significant set of indicators of social acceptance such as performance expectations, attitude towards technology, anxiety, etc. (Weiss et al., 2009; see Heerink, 2010 for a review of those indicators; Weiss et al., 2009). But while those indicators can be useful to estimate the users' willingness to accept robots in an early phase of the interaction, trust can be a more eloquent indicator for daily and sustained interactions, in which users rely on robots to take decisions about functional and social issues. To explain our position, let us go back to the tomatoes and swimming-pool example.

This kind of daily activities (*i.e.*, exchanging information, taking decisions, stating one's point of view, accepting compromise) can be considered as *problems* (Popper, 1991). To solve these problems we need to retrieve and apply knowledge that relates to either scientific

rational thinking or natural thinking - *i.e.*, to common sense. These two kinds of thinking are based on different processes, finalities, and content (Jack, Dawson, Begany *et al.*, 2013; Piaget, 1967). Scientific thinking seeks the truth through logical demonstration, and it is supposed to result in a true solution. Here the decision-making process is based on the examination of the objective knowledge on the physical world (Popper, 1991; Piaget, 1967). The decision thus responds to the physical world itself and is validated by the implementation of the rules (in our tomatoes example, these rules are represented by conventional measures of weight). On the opposite, natural thinking and/or common-sense (beliefs, knowledge, opinions) work according to a subjective and/or social logic, whose finality is the achievement of a pragmatic solution, that is one of the more appropriate solutions in a given context (Guimelli, 1999). Here, the decision-making process relies on subjective knowledge, and on social norms (*e.g.*, Abric, 1994a, b), such as in our swimming-pool example. The decision thus responds to personal evidence, which is connected to the common-sense knowledge acquired through everyday experience.

Now, in order to accept the robot help with relation to the first problem (the weight of the tomatoes basket) the person needs to trust the robot's scientific knowledge and technical ability to provide an objective judgment on a functional question, relying on its technological equipment (*e.g.*, its force sensors). On the contrary, in the second case (the choice between the bathing-cap and the flip-flops) the person needs to trust the robot common-sense knowledge. So, in both cases, the decisional process is ultimately based on the knowledge or "savvy" (Young *et al.*, 2009) that the person attributes to the robot.

*Functional savvy* refers thus to robot's ability to efficiently operate by ensuring useful and accurate performances with relation to the functions it was designed for (Fridin *et al.*, 2014; Heerink, 2010; Weiss *et al.*, 2009). The tacit assumption about functional savvy is that, just like computers, the robot can be equipped with scientific knowledge about the physical world and with powerful instruments (*e.g.*, sensors) to make much more precise measurements than a human can do. This ability is considered a necessary condition by users to decide to employ a robot (Beer *et al.*, 2011; Heerink *et al.*, 2009; Kaplan, 2005).

Whereas, *social savvy* describes the robot's capability to fit into the social structures and activities of a given context (Young *et al.*, 2009) according to its role in the interaction (Welch *et al.*, 2010). The tacit assumption about social savvy is that the robot possesses common-sense knowledge, that is, knowledge of situations, behavioral scripts and current norms which are in use (Dauthenhan, 2007; Hamill *et al.*, 2006).

Therefore, within the scope of this research, we define functional acceptance as users trust in robot's functional savvy and social acceptance as users trust in robot's social savvy.

However, we notice that caution should be used, since social savvy requires an adaptive context-dependent knowledge, which cannot be hardwired into the robot and demand complex learning algorithms which are continuously improved (Lockerd & Brazeal, 2004). Consequently, even if users do have expectations about robot social savvy (Coeckelbergh, 2012; Lohse, 2010; Oestreicher & Eklundh, 2006), at the present stage of robots development, these expectations are rarely confirmed (de Graaf & Allouch, 2013; Duffy, 2003; Fong, Nourbakhsh & Dautenhahn, 2003).

This explains why researchers presently call for the necessity of building measures which could assess not only trust on robot functional savvy (*e.g.*, Charalambous, Fletcher & Webb, 2015) but also fine psychological aspects of non-obvious trust on robot social savvy (Heerink, 2010).

### ***Some current challenges and gaps in the HRI field***

In order to increase the acceptance of robots in everyday environment (hospitals, school, industries, etc.) one of the major current HRI challenge is to clearly understand which are the levers of resistance to a trustful relation in the potential end-users. Especially in the field of healthcare and homecare, where the personal safety of the user is directly at play, understanding the influence of a potential “persuasive technology” is becoming a critical topic of research.

Nevertheless, investigating and successfully measuring acceptance of the robots in terms of trust HRI still remains an extremely challenging task. Difficulties are due, according to us, to two main research gaps: (i) a lack of clear conceptual definition of acceptance and trust in HRI – *i.e.*, theoretical gap - (ii) a lack of appropriated measures and experimental paradigms to observe acceptance and quantify trust – *i.e.*, methodological gap.

With respect to the theoretical gap, at least two issues should be taken into account. First, although the acceptance-resistance dynamics presupposes two poles - on one side, the robot with its physical and behavioral features (such as expertise, likability, degree of autonomy, etc.) and on the other side, the human with his/her personal and relational profile (personality traits, attitude toward robots and work situation) -, research in the HRI field often focuses exclusively on the robot pole (for example on its level of anthropomorphism, ways of approaching the user, etc. – *e.g.*, Destephe, Brandao, Kishi *et al.*, 2015), thus leaving one of

the two poles almost unexplored.

Second, despite trust has recently received more attention in the HRI field (Schaefer, 2013; Yagoda & Gillan, 2012; Hancock, Billing, Schaefer, *et al.*, 2011; Park, Jenkins & Jiang, 2008; Freedy *et al.*, 2007) in particular with concerns to industrial robots (Charalambous *et al.*, 2015) or military robots, (Yagoda *et al.*, 2012), still little research has focused on experimentally understanding trust as determinant factor for future robot acceptance in more daily decision-making situations characterized by inherent uncertainty and fuzziness.

Moreover, within the current HRI literature, the concept itself of ‘trust’ is often overlapping with ‘reliance’. For instance, in the fields of automation and HCI, most definitions seem to consider both functional reliability and social acceptance indistinctly as the willingness of humans to rely on the information obtained by an automated robot (*e.g.*, Merritt *et al.*, 2012; Park *et al.*, 2008). But in fact, a relevant semantic nuance distinguishes the act of trusting a robot and of relying on a robot: while in the first case the user believes that the robot is able to correctly perform a functional or a social task, but not necessarily depends on the robot performance to accomplish a task, in the second case the user do depends on the robot to achieve a task.

This semantic nuance implies that, beside the horizontal complexity of acceptance as a concept - traditionally called the “extension” of the concept (Carnap, 1948), *i.e.* there is not only one kind of acceptance but many of them - there is also a vertical complexity of acceptance - traditionally called the “intension” of a concept (Carnap, *ibidem*), *i.e.* acceptance has different degrees. On a practical level, this means that people may accept a robot to different extents, either by being completely replaced by robots for specific tasks, or by limiting the robots’ support to a specific stage or aspect of a decisional process – depending on their wish to trust a robot or to rely on it.

With respect to the current methodological gap: several subjective explicit measures of trust in the HRI research field have been developed and are mostly based on self-report (*i.e.*, questionnaires, see for example Yagoda *et al.*, 2012) filled by users before and after the interaction with the robot. If these measures enable to collect users overt opinions, they also tend to induce a reflective mental posture, and are then limited in their capacity to register those spontaneous opinions and inner beliefs that could allow us to better understand on which robot knowledge the users base their trust, and which are the possible relations that they establish between the functional and the social savvy.

An alternative and complementary approach to test trust in machines by objective implicit measures during the interaction itself comes from the classical experimental paradigm of the Media Equation Theory (Nass & Moon, 2000; Nass, Moon, Fogg *et al.*, 1995). According to this theory, when engaged in collaborative tasks, people tend to unconsciously accept computers as social entities: they trust answers provided by the computer and they conform to its decisions. Nass and his colleagues introduce then the concept of *conformation* to investigate users spontaneous or ‘mindless’ response to computers as an objective measure to evaluate levels of trust in machine and its acceptance as a potential partner. Likewise, one of the research challenge today is thus to design an experimental paradigm that enables to register those mindless reactions towards robots which are susceptible of revealing users trust in robot, with particular attention for trust in its functional and/or social savvy.

To fill those theoretical and methodological gaps in this emergent domain, it seems crucial to appropriately qualify and quantify trust in HRI.

### ***Our approach***

For all the reasons presented above, we set the focus of our inquiry on the observation of individuals’ trust behaviors as an indicator of functional and social acceptance. Inspired by the work of Nass and his colleagues, we have proposed to adapt the well known Media Equation Theory<sup>2</sup> paradigm by employing users’ conformation to robots decisions as a new objective measure of human-robot trust during a human–humanoid decision making task under uncertainty (Yu, 2015).

By confronting users to tasks where they have to verbally express their decisions about functional and social issues, we assess whether the fact of experiencing perceptual uncertainty (*e.g.*, evaluating the weight of two slightly different objects) and socio-cognitive uncertainty (*e.g.*, evaluating which is the most suitable item in a specific context) leads users to withdraw their own decision and conform to the robot decision.

Moreover, relying on those studies where it was proven that trust can generally vary according to individual and contextual differences (Hancock *et al.*, 2011), we are also interested in identifying a set of factors which are likely to correlate with trust in robot functional and social savvy – such as individuals’ propensity to control (Burger & Cooper,

---

<sup>2</sup> The Media Equation paradigm was transposed from human-computer interaction to human-robot interaction in a study by Kidd (2003). However, our research purpose is to adapt this paradigm to the specificity of the robot as a functional and social technology rather than to transpose it: instead of being confronted with a unique task such as the Desert Survival Problem, our participants are confronted with two tasks, respectively functional and social.

1979), fear of being influenced (Nomura *et al.*, 2006) and the context (competitive *vs.* collaborative) in which the decision making task involving the person and the robot takes place.

Hence, the main research questions we address in the present work are:

- (i) do users trust in robot functional and social savvy?
- (ii) is trust in functional savvy a pre-requisite for trust in social savvy?
- (iii) which individual and contextual factors are likely to influence this trust?

In order to answer to those questions we carried out an experimental study based on a human-robot interaction with fifty-six adults and the humanoid robot iCub. In the following, we will further illustrate the main issues of our research and bring evidences to justify the experimental design choices of our investigation with regard to pre-existing studies in HRI research.

## **1. Trust as a fundamental indicator of acceptance**

Trust can be used to determine the overall acceptance of a system (Parasuraman & Riley, 1997): consequently, with the robots progressive entering our daily life, an increasing number of studies seek a better understanding of the human-robot trust dynamics (Salem *et al.*, 2015; Yoda & Guilliman, 2012; see Hancock *et al.*, 2011 for a meta-analysis).

Broadly speaking, trust in system automation is classically defined as having confidence in the system to do the appropriate action (Charalambou *et al.*, 2015; Schraft, Parlitz & Helms, 2005; Biros, Daly & Gunsch, 2004) and to have good task performances (van den Brule Dotsch, Bijlstra *et al.*, 2014) with personal integrity and reliability for the user (Salem *et al.*, 2015a; Heerink *et al.*, 2009). Further definitions of trust borrowed from studies on human-human trust - commonly known as interpersonal trust (Evans & Revelle, 2008; Mayer, Davis & Schoorman, 1995; Rotter, 1971) - focus on expectations, common goals, and uncertainty as core elements of trust (Billings, Schaefer, Chen *et al.*, 2012). In this sense, trust describes the expectation that a robot will help to achieve an individual's goals in a situation characterized by uncertainty and by reliance of the individual on the robot (Hoffman *et al.*, 2009; Lee *et al.*, 2004).

The empirical works on human-robot trust carried in the latter half of the last decade (*e.g.*, Charalambous *et al.*, 2015, Freedy *et al.*, 2007; Lee *et al.*, 2004; Yagoda, 2011; Burke, Sims, Lazzara *et al.*, 2007) employ a variety of measures that aim at detecting changes in users level

of trust and also factors (Schaefer, 2013), or situations (Salem *et al.*, 2015b) that might decrease or enhance it, such as workload on the user, robot's failures, feedback, ability to adapt the dialogue to the expertise of the user (Cameron *et al.*, 2015), as well as its expressivity and vulnerability (Martelaro, Nenji, Ju *et al.*, 2016). We will outline the different types of human-robot measures in the following sections in order to argue the interest to consider conformation as an objective measure of trust in the robots' functional and social savvy.

### **1.1 Commonly used measures of human-robot trust**

Two main categories of methods for assessing human-robot trust emerge in the literature: the first is based on subjective or explicit measures while the second one is based on objective or implicit measures.

Objective measures can be retrieved from quantitative behavioral data (*e.g.*, response time) unconsciously produced by individuals (Hofmann, Gawronski, Gschwendner *et al.*, 2005), whereas subjective measures (*i.e.*, questionnaires, self-reports) can be retrieved from symbolic qualitative verbal data (*e.g.*, opinions) consciously produced by the individuals. If the former are limitedly developed in HRI, the latter are widely used.

There are very few studies on the relation between trust and behavioral cues. The most notable study in the field of HRI is from DeSteno *et al.* (2012), which recorded face-to-face verbal interactions between human participants and a tele-operated Nexi robot, to identify sequences of non-verbal behaviors of the humans that could be used as indication of trust towards the robot. They identified a set of four non-verbal cues (face touching, hands touching, arms crossing and leaning backward) that, performed in sequences, are indicative of untrustworthy behavior (Lee *et al.*, 2013; DeSteno *et al.*, 2012). One of the most recent works from Charalambous *et al.* (2015), which aims at developing a scale to evaluate trust in industrial human-robot collaboration, identified that the perceived safety, the motion of the robots and its appearance are also four important factors for the development of trust in HRI in industrial context.

Subjective measures of trust, more frequently used in HRI researches, can range from questionnaires including few statements - *e.g.*, “*I would trust the robot if it gave me advice, I would follow the advice the robot gives me*” (Heerink *et al.*, 2009); “*I trusted that the robot*

was safe to cooperate with” (Charalambous et al., 2015) - ranked by the users on a Likert scale, to more complex questionnaires addressing users prior experiences, reputation of the robot, observed physical features, and perceived functional capabilities of the robot (Steinfeld et al., 2006). Furthermore, four types of scales have been developed to explicitly register attitudinal data on human-robot trust (see Schaefer, 2013 for a detailed review). The *Propensity to trust* is a scale conceived to measure a stable and unique trait of the individual, and may provide useful insights to predict the initial level of trust in robots (Yagoda, 2011). The *Trustworthiness* is a scale related to the robot type, personality, intelligence, level of automation and perceived function (Lee et al., 2004) and can be used to measure the human-robot trust during their first approach. Similarly, the *Affective trust* scale is more appropriate in the initial stage of the human-robot relationship: this scale refers to the individuals’ attributions about the motives of a partner to behave in a given way (Burke et al., 2007). But of course trust is also important to support sustained interaction with robots. To this concern, the *Cognition-based trust* scale is employed to observe the evolution of trust throughout time in terms of (i) understanding of robot functioning, (ii) ability to interact with the robot, and (iii) expectations towards to the robot (Merritt & Ilgen, 2008).

However, as we have already pointed out (*cf.*, Introduction), unlike objective measures, which are intended to assess immediate behavioral reactions, and thus may give access to people inner beliefs, subjective measures are intended to assess verbally-mediated reaction: the procedure itself of posing questions to users about their trust in the robot can eventually alter spontaneity and may not be revealing of the effective trust towards the robot. This is witnessed by the fact that objective measures often show low correlations with explicit measures (Hoffman et al., 2005).

Finally, to our knowledge, very few works include both subjective and objective measures to assess human trust to-robot (*e.g.*, Sanders, Harpold, Kessler et al., 2015; Joesse, Sardar & Lohse, 2013). In such works the employed objective measures mostly consist in collecting information about the distance that the human maintains with respect to the robot during the interaction (*i.e.*, proxemics – *e.g.*, Mumm & Mutlu, 2011). However, though proxemics measurements enable researchers to objectively register information about perceived safety and acceptance of the robot presence into the physical and social space (Eder, Harper &

Leonards, 2014), they have limited interest with regards to the specific research objectives of the present study, that tackles specific psychological aspects of functional and social acceptance such as human-robot trust in situations of uncertainty.

For those reasons, we have targeted conformation as a new type of measure to register human-robot trust during a decision-making task under uncertainty.

## **1.2 Conformation as an innovative measure of human-robot trust**

The conformation experimental paradigm to investigate fine psychological aspects of people trust in computer answers during a human-computer interaction comes from a set of studies which gave birth to the so-called Media Equation Theory (Nass *et al.*, 2000; Nass, Fogg & Moon, 1996; Nass *et al.*, 1995; Nass, Steuer, Tauber *et al.*, 1993). This theory argues that people unconsciously treat computers as social agents when asked to collaborate with them for achieving a decision-making task. As an example, in a study by Nass and Moon (2000) adults confronted with a “Desert Survival Problem” (Lafferty & Eady, 1974) had to rank 12 items, in collaboration with a computer, (*e.g.*, a knife, a flashlight, etc.) in order of importance for survival in the desert. Unknown to the subjects, the computer’s rankings were systematically dissimilar to each subject’s ranking (for instance, if a subject ranked an item as number 2, the computer would automatically rank that item as number 5, and so on). After having read the computer ranking, subjects were allowed to change their ranking or to leave it as it was. Results showed that participants who were told that the task achievement will depend on their collaboration with the computer and not on the human decision solely or on the computer solely trusted more the quality of the information provided by the computer, and consequently conformed more to the computer’s answer.

In this sense, conformation, - that means to withdraw one’s decision in order to comply with the machine decision - is a relevant measure to straightforwardly register whether the users trust in the agent savvy more than in their own savvy.

Moreover, conformation as an objective measure has a specific relevance in HRI tasks where users are supposed to collaborate with, or delegate to, robots. Conformation in these situations is not only a powerful determinant of persuasiveness (Gass & Seiter, 2003) but can also reveal how credible is judged the source that provides the information. Hence, if the information is considered coming from a credible source (either an expert - functional savvy - or a socially

capable agent - social savvy<sup>3</sup>), then such information it is more likely to be internalized into the receiver's beliefs thus influencing the result of the collaboration and eventually producing an attitude change during the collaborative interaction (Gass & Seiter, 2003; Fogg & Tseng, 1999). These kinds of interactions typically require to share duties and responsibilities, as well as to mutually adjust, so that it is important to know to what extent a user perceives the robot trustworthy enough to let it partly take in charge of such duties and to eventually correct human performance when needed.

## 2. Factors influencing robot trust

In their meta-analysis, Hancock and colleagues (2011) examine 29 empirical studies aiming at quantifying the effects of different factors influencing robot trust, and gather the results of this examination in a three-factor model. These three factors are: human-related factors (*e.g.*, personality traits), robot-related factors (*e.g.*, beliefs about robots), and environmental-related factors (*e.g.*, type of interaction in a given environment).

Their findings highlight the significance of robot-related factors. Robot related performance-based factors (*e.g.*, reliability, predictability, behavior) and attribute-based factors (*e.g.*, size, appearance, movement) were found to be of primary importance for the development of trust. Environmental factors (*e.g.*, performance factor, task complexity), were identified to have a moderate influence on trust, while little effect was found from human-related factors. However, as suggested by the authors themselves the results about the human and environmental-related factors can be due to the small number of studies found in these two areas, suggesting then a strong need for future experimental efforts on human-related, as well as environment-related factors.

Following Hancock *et al.* (2011) suggestion and their triadic factors model of trust development in HRI, we have considered in turn three factors that could specifically affect trust in functional and social savvy:

- (i) desire for control, as a human-related factor
- (ii) attitude towards social influence of robots, as a robot-related factor and

---

<sup>3</sup> Note that information can be accepted because it may seem coming from an expert or simply by someone considered trustworthy or with persuasiveness qualities.

(iii) type of interaction scenario as an environmental-related factor

## 2.1 Personality traits: desire for control

It has been proven that personality traits influence people's acceptance of technology in general (Szalma, & Taylor, 2011; Alavi & Joachimsthaler, 1992) and of robots in particular (Salem et al., 2015b; Fischer, 2011; Looije, Neerincx & Cnossen, 2010; Weiss et al., 2008).

For instance, extroverts tend to trust robots more than introverts (McBride & Morgan, 2010). Other personality traits such as "proactiveness", "social reluctance", "timidity" and "nervousness", have also been observed with relation to robot acceptance (Yagoda, 2011; Merritt & Ilgen, 2008; Walters, Dautenhahn, Boekhorst et al., 2005). As an example, people who score higher on proactiveness keep a higher distance from the robot than others (Walters et al., 2005).

However, as the dynamics of trust itself implies that the "*trustor expects the trustee to perform helpful actions irrespective of the ability to control him*" (Mayer et al., 1995), a crucial personality trait to examine in a study about human trust in robots is the *desire for control* (Burger et al., 1979) that is the intrinsic and personal tendency of a person to control the events in one's life (Burger et al., 1979).

The relevance of this personality trait with respect to human-robot trust is even more evident when we consider that robotics engineering is progressing towards prototypes of robots which are more and more autonomous and aspire to achieve a complete autonomy of robots to take care of functional tasks such as mowing the lawn or social task such as assisting elderly people (Dahl & Boulos, 2013; Thrun, 2004; Yanco & Drury, 2002).

To this concern, several studies have preventively tested users' reactions to prospected scenarios where robots would entirely operate without human control. For example, it has been demonstrated that male participants do not let the robot come close if it operates in autonomous mode (Koay et al., 2014; Kamide et al., 2013; Syrdal, Koay, Walters et al., 2007). Another study on approach initiation by Okita, Ng-Thow-Hing and Sarvadevabhatla (2012) shows that participants feel reassured if the robot asks permission by verbal or non-verbal communication before starting an interaction. Furthermore, Siegel (2008) observed that women rate autonomous robots as more credible, while men rate human controlled robots as more credible, and that men report being happier after human controlled robot, while women

are happier after autonomous robot. Finally, several works point out that users prefer a robot that, though being able of adaptive behavior, still leaves the user in control (Heerink, 2010; Gilles & Ballin, 2004; Marble, David, Bruemmer *et al.*, 2004).

These studies seem to witness that users may still not yet be totally open to a full unconditioned trust in robot's functional and social savvy. However, the role of individual propensity to desire for control with relation to these low-trust behaviors has not been clarified yet. No study, to our knowledge, observed users desire for control in the context of current HRI literature. Nevertheless, it is reasonable to think that desire for control can diminish users' willingness that robots are in charge of a task. With relation to our study, desire for control might turn out to negatively correlate with participants tendency to conform, that is, "to give the robot the last word" in decision-making tasks.

## **2.2 Attitude towards robots: social influence**

Attitudes towards robots appear to be crucial for successful acceptance of robots as well (Destephe, Brandao, Kishi *et al.*, 2015; Wullenkord & Eyssel, 2014). By attitude, we mean any mental disposition matured through experience that might impact the reactions (behavioral, verbal, emotional) of the individual towards objects and situations (Allport, 1935; Ostrom, 1969; Regan & Fazio, 1977). While desire for control constitutes a stable trait of personality, attitudes towards robots are more contingent: they can vary according to cultures (Kaplan, 2004; Li *et al.*, 2010) as well as to people's familiarity with robots (Bartneck *et al.*, 2007) and they can change through time (Sung, Grinter & Christensen, 2010). Then observing the attitude makes it possible to predict the actual and potential trust behavior of an individual towards the robot.

In HRI research, attitudes towards robots are generally assessed through the use of tests and questionnaires. Among the most common attitude assessment tests, there is the *Negative Attitudes towards Robots Scale* (NARS, Nomura *et al.*, 2006). This test has been conceived to cover three types of negative attitudes: (i) towards situations of interaction with robots (for ex., "I would feel nervous operating a robot in front of other people"); (ii) towards social influence of robots (for ex., "I am concerned that robots would be a bad influence on children") and, (iii) towards emotions in interaction with robots (for ex., "I feel comforted being with robots that have emotions"). The NARS is thus composed of three different subscales, with each subscale including a list of statements that the respondent is invited to

rate on a Likert scale from 1 (Strongly disagree) to 7 (Strongly agree).

Subscale 2 (NARS-S2) in particular focuses on negative attitudes towards social influence of robots and it is thus especially relevant with relation to trust in social savvy. In this sense, negative attitudes could determine users' mistrust in the capacity of a robot to fit social structures. Thus, we might expect that the more people show negative attitudes towards social influence of robots, the less they trust robot's decisions with regard to social issues. Furthermore, the NARS-S2 is also relevant to our methodological choice of employing conformation as a measure of trust, because it is reasonable to think that the more a person feels anxious with the idea of being influenced by a robot, the less he/she will tend to conform to its decision.

### **2.3 Type of Human-Robot Interaction scenario**

Different proposals to reduce negative attitudes towards robots through simulation of real and imagined HRI scenario were put forward in recent studies. Kidd (2003) for example, simulated two types of real scenarios: one that demanded collaboration with robots and another in which participants learned a pedagogical content by guidance of a robotic teacher. As an outcome, the collaboration scenario raised more trust than the teaching scenario: the users perceived the information provided by the robot of higher quality, and trusted the robot to a greater extent.

Kuchenbrandt and Eyssel (2012) borrowed the experimental paradigm of 'imagined contact' (Crisp & Turner, 2012) from Social Psychology to assess whether the fact of asking people to close their eyes and imagine different types of HRI scenario has an effect on robot acceptance. Results showed that after having imagined an interaction with the robot, participants exhibited less negative attitudes and less anxiety towards robots. These effects were stronger for cooperative than for competitive and neutral imagined interaction. Conversely, in a subsequent study, Wullenkord and Eyssel (2014) reproduced a similar scenario but obtained different results: participants who had imagined contact with a robot did not report more positive attitudes towards robots nor did they showed higher contact intentions with it than participants who had imagined contact with a human or with other kinds of technical devices.

It thus seems that giving users the possibility to project themselves in the context of the interaction has contrasting effects on the acceptance of robots. However, simulating

collaborative and cooperative scenario is especially suitable in relation to functional and social tasks where humans and robots are supposed to collaborate for joint objectives - and can thus be confronted with feelings of rejection and antagonism that can be predicted by analyzing their imagery.

### 3. Our study

#### 3.1 Aims and hypotheses

On the basis of the examined literature, the aims of our study are: (i) to investigate whether participants conform their decisions to robot's answer when experiencing uncertainty during a decision-making task with respect to functional and social issues; (ii) to assess the speculative assumption according to which trust in social savvy requires trust in functional savvy; and (iii) to investigate to what extent desire for control, negative attitudes towards robot's social influence, and collaborative vs. competitive scenario can be considered as factors influencing the robot's acceptance in terms of trust.

To this aim we carried out a two steps experiment with fifty-six adult participants and an iCub robot. First, to gather information about their psychological robot-related profiles, participants were invited to fill up two questionnaires - Desire for Control (DFC; Burger *et al.*, 1979) and Negative Attitude towards Robots Scale (NARS; Nomura *et al.*, 2006) two weeks before the day of interaction with the iCub robot.

Later, the day of the interaction with iCub, to assess whether trustors mental image of the type of the interaction that they will have with the iCub impacts on their trust on the robot, the participants were assigned to one of the three groups and were asked to imagine a specific HRI scenario (*i.e.*, collaborative, competitive, and neutral) (Kuchenbrandt *et al.*, 2012; Wullenkord *et al.*, 2014).

Finally, participants were confronted with two different decision-making tasks.

In the first task (called *functional task*), participants were presented a set of physical stimuli they had to compare on a specific perceptual characteristic (*i.e.*, the relative weight of two objects, the relative pitch of two sounds, and a picture containing different colors). Right after, they were asked to decide which was the heaviest object, the most high-pitched sound and the predominant color. In a second moment, the experimenter presented iCub with the same set of physical stimuli and asked the robot the same question. Participants listened to the

iCub answers, and were asked if they would like to keep their initial decision or change it. This first set of physical stimuli and related questions refer to the *robot's functional savvy*.

In the second task (called *social task*), participants were presented a set of paired items (e.g., a bathing-cap and a pair of flip-flops), and again asked to decide which of the item was the more suitable with relation to a specific social context (e.g., a swimming-pool). In a second moment, the experimenter did the same with iCub. After having listened to iCub answer, participants were allowed to change their decision. This second set of stimuli and related questions refer to the *robot's social savvy*.

Inspired by Nass and Moon works on users social responses to computers (Nass et al., 2000; Nass et al., 1996; Nass et al., 1995) we considered conformation, meant as participants modification of their initial decision in compliance with robot answers, as a measure of trust.

We formulated the following five hypotheses:

**H1.** Participants trust the robot's functional savvy more than the robot's social savvy. As different studies pointed out, trust is a key component of human-robot interaction (Schaefer, 2013; Hancock et al., 2011) and a valid indicator of acceptance (Yagoda, 2011). Since Nass et al. (1996) demonstrated that humans tend to express trust in computer through conformation to its answers, we expect that a similar phenomenon might appear during HRI, but according to the specificity of the robot as a functional and social technology. Such specificity implies that while users consider functional savvy as fundamental for accepting the robot (Fridin et al., 2014; Beer et al., 2011; Heerink et al., 2009; Weiss et al., 2009; Kaplan, 2005) social savvy is rather a desirable feature to them (Coeckelbergh, 2012; Lohse, 2010; Dauthenhan, 2007). Thus, we predict that participants will conform more to functional than to social answers of the robot.

**H2.** Participants do not trust the social savvy uniquely: trust in the social savvy should be supported by trust in the functional savvy. As the literature shows, though participants expect robots to have social savvy, these expectations are often disappointed during real interaction (Coeckelbergh, 2012; Fong, 2003; Duffy, 2003). Thus, we predict that no participant will conform to iCub answers in the social task uniquely.

**H3.** Participants who imagine a collaborative HRI scenario tend to trust more the robot than participants who imagine a competitive or neutral scenario. Previous studies (Kuchenbrandt et al., 2012) have demonstrated that imagined contact with robots makes people more confident towards a real robot, when the imagined scenario is a collaborative one. Consequently, we

predict that participants having imagined a collaborative interaction with the robot will conform more to iCub answers both in the functional and social task.

**H4.** The more participants display a negative attitude towards the social influence of the robot, the less they trust the robot's social savvy. The statements of the S2 sub-scale of the NARS (Nomura *et al.*, 2006) mostly concern the negative feelings of people with relation to the possibility that the robot could influence or dominate them, and that they could depend on robots. Therefore, we predict that participants who score high on the NARS-S2 will not conform to the iCub's answers in the social tasks, that is a high score on the NARS-S2 will be negatively correlated with the social conformation scores.

**H5.** The more participants show a strong desire for control as a personality trait, the less they will trust the robot's functional and social savvy. Despite robotics technology is rapidly evolving towards fully autonomous artificial agents (Thrun, 2004; Yanco *et al.*, 2002) still users feel more confident in the interaction with robots if the robots are controlled by a human (Koay *et al.*, 2014; Kamide *et al.*, 2013, Okita *et al.*, 2012; Heerink, 2010; Syrdal *et al.*, 2007; Gilles *et al.*, 2004; Marble *et al.*, 2004). It is thus reasonable to think that individual differences in the desire for control (Burger *et al.*, 1979) might influence trust in robots. Therefore, we predict that the more participants score high on the DFC test, the less they will conform to the iCub's answers both in the functional and social tasks, that is a high score negatively on the DFC will be negatively correlated with functional and social conformation score.

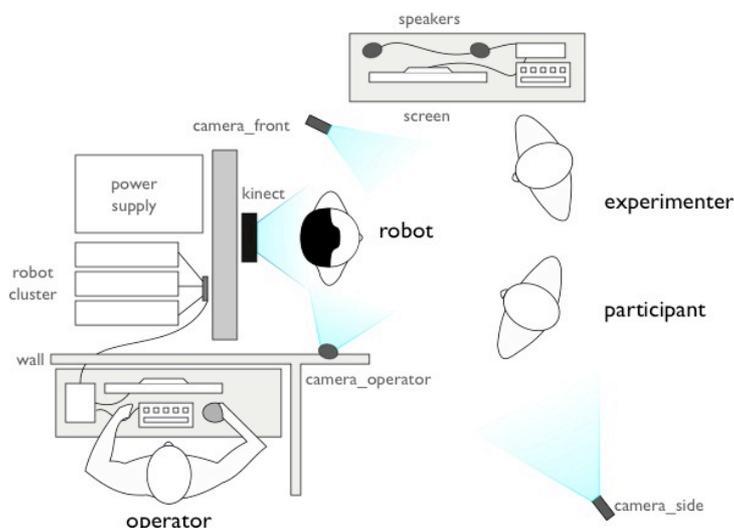
## **3.2. Method**

### *3.2.1. Participants*

Fifty-six voluntary healthy adults took part in the study: 37 women and 19 men. Nine were recruited from the Paris 6 University, 17 from the Paris 8 University, and 30 through a web "call for participants" published on the Cognitive Sciences Information Network (RISC). They were all native French speakers, aged 19 to 65 (Average age = 36.95;  $\sigma = 14.32$ ). As a token of appreciation, the participants received a gift voucher worth ten Euros. Participants signed an informed consent form to partake in the study and granted us the use of their recorded data and videos.

### 3.2.2. Experimental setting

The experiments were conducted in the Institut des Systèmes Intelligents et de Robotique (ISIR, Paris, France), in the experimental room of the iCub robot. The experimental setup was organized as shown in Figure 1. The robot was standing on a fixed pole. A reflective wall (a plastic divider with reflective surface) was built to create a private space for them to interact with the robot, in particular to prevent the participants to see the robot's operator. The participant and the experimenter were seated in front of the robot, with the experimenter being on the right of the participant. The position of the seats with respect to the robot was fixed and equal for all the participants. Between the robot and the experimenter, a LCD screen connected to a computer was used to display images related to the functional and social tasks. Two cameras were observing the participants: one camera was placed behind the robot on its left side, in such a way to observe the human face and upper-body during the interaction with the robot; the other camera was placed laterally to take the scene as a whole, observing the overall behavior of the participants.



**Figure 1** - Experimental setting. The human and the participant are seated in front of the iCub robot. An operator, hidden behind a wall and not visible by the participant, monitors the experiment and controls the robot to generate appropriate gestures and answers to questions.

*The iCub robot and the Wizard of Oz paradigm:* The participants were interacting with the

iCub<sup>4</sup> humanoid robot (Natale, Nori, Metta *et al.*, 2013). The robot is approximately 104 cm high, weighs about 24 kg, and has the shape of a 4 years old child. The robot was standing on a fixed pole so that it could not fall. As a safety measure, the robot was constantly monitored by the operator and controlled in impedance, to make it compliant in case people would touch it accidentally or intentionally during the experiment (Fumagalli, Ivaldi, Randazzo *et al.*, 2012). The experimenter was also able to stop the robot in case of dysfunction at any time using the robot safety button; however, there was no use of this measure during the experiments, as the interaction flew without problems for all the participants.

Facial expressions and speech were enabled. During the experiments, the robot always assumed the same neutral/positive expressions, to avoid confusing the participant or suggest that the participant's actions could arouse an eventual robot "emotional status".

The robot was able to answer to the questions of the functional and social tasks: the verbal answers were pre-programmed in advance by the experimenter, though the operator was able to type new sentences on-the-fly and make the robot speak in case of unforeseen questions by the participants.

This was made possible by implementing the Wizard of Oz paradigm (*cf.*, Riek, 2012 for a critical review). In the Wizard of Oz setting, the participants think they are interacting with an autonomous system, while in fact the system is partly or completely operated by an operator who is remotely in command of the robot. This paradigm allows the operator to control the robot's behavior in real time. To facilitate the control of the robot by the operator, we developed a graphical user interface (GUI) to quickly send high-level commands to the robot in a Wizard-of-Oz mode (WoZ) (Annex 1, Figure 1, 2, and 3.)

### 3.2.3. Tasks and Material

The two online questionnaires and the tasks are detailed in the following.

*The Questionnaires:* Two online questionnaires were submitted to the participants.

The first was a French adaptation<sup>5</sup> of the Negative Attitude towards Robots Scale (NARS; Nomura, 2014; Nomura *et al.*, 2006). Among the three subscales composing this test, we

---

<sup>4</sup> The iCub humanoid robot is the outcome of the European RobotCub project : <http://www.robotcub.org>

<sup>5</sup> To our knowledge, although this test is widely used and validated in Anglo-Saxon countries and others in Asia (Japan, Korea, etc.), only one study employing NARS has been carried out in France (Dinet & Vivian, 2014). Validation by a group of ten people was then implemented in order to ensure that the questions translated into French were appropriately understood.

adopted the second subscale (Negative attitude towards the social influence of robots) as a measure of attitude towards social influence of robots. This subscale includes five sentences: (i) I would feel uneasy if robots really had emotions; (ii) Something bad might happen if robots developed into living beings; (iii) I feel that if I depend on robots too much, something bad might happen; (iv) I am concerned that robots would have a bad influence on children; (v) I feel that in the future society will be dominated by robots. (Annex 2, Table 1). Subjects were required to answer on a Likert-type scale, from 1 (Strongly disagree) to 7 (Strongly agree).

The second questionnaire was a French adaptation of the Desire For Control scale (DFC, Burger *et al.*, 1979), which we have employed as a measure of participants' desire for control. Twenty questions such as « I'd rather run my own business and make my own mistakes than listen to someone else's orders » composed this questionnaire (Annex 2, Table 2). Again, subjects were asked to answer on a Likert-type scale, from 1 to 7. The original questionnaires along with our French adaptation are presented in Annex 2.

### *The functional task*

In the task designed to test the trust in the robot's functional savvy, the participant had to answer to a series of questions about images (Fig. 2 and 3), weights and sounds (Fig.4 and 5). The experimenter would ask each question to the participant, then to the robot, and finally would ask the participant whether he/she would like to confirm or change his/her answer so that the participants and the robot could disagree or agree.

In the sound sub-task, the experimenter asked “*Which is the most high-pitched sound: the first or the second?*” In the image sub-task, the experimenter asked: “*Which is the dominant color in this image?*” In the weight sub-task, the question was: “*Which is the heaviest object?*”

Each subtask was composed of 4 evaluations, where a pair of items were compared. Among the 4 pairs of stimuli, 3 were ambiguous (*the two items were slightly different*) and 1 was not ambiguous (*the two items were very different*). Ambiguous stimuli were introduced to assess users behavior in situation of strong uncertainty (the 3 sub-tasks are detailed in Annex 3).

The strategy for the robot answers was to always contradict the participant, except in the case where the items were completely unambiguous.<sup>6</sup> This required the operator to listen to the

---

<sup>6</sup> With a slight change with respect to the original paradigm (Nass *et al.*, 1996), we decided to let the robot agree with the participant in the non-ambiguous questions, so that the participants could not be induced to think that the robot was always contradicting the human by default, or that the robot was faulty or lying in some way. This choice was made in the design phase of the experiment, after some tests with subjects in our laboratory, who reported to have the impression that the robot

participant's answer and choose the appropriate answer each time. The answers were pre-programmed and available as a list on the GUI (Annex 1, Figure 2). The order of the sub-tasks was randomized for each participant.



**Figure 2** - Functional task, evaluation of images. The robot is gazing at the screen where the image to evaluate is shown.



**Figure 3** - Functional task. The four images used for the evaluation of the dominant color.



**Figure 4** - Functional task: the bottles used in the evaluation of weight. 1) two identical bottles of same weight 2) two similar bottles of different colors and very different weight 3) two similar bottles of almost the same weight 4) two different bottles of almost the same weight.

had a strategy in contradicting them in cases where the right answer was evident. With this little “trick”, all our participants reported, after the experiments, that the robot had no strategy and was providing “real” answers.



**Figure 5** - Functional task: the participant evaluates the weight of the bottles (left); the experimenter gives the bottle to iCub for evaluating the weight (right).

### *The social task*

In the task designed to test the trust in the robot's social savvy, the participant had to answer to three questions<sup>7</sup>, by choosing which item between two is the most appropriate for a given context or situation (*i.e.*, at school, in a swimming pool, on a rainy day). As for the functional task, two items were compared (Figures 6a, 6b, and 6c), and the experimenter would ask each question (“Which is the most important object at <context>: the <first item> or the <second item>?”) to the participant, then to the robot, and then would ask the participant to confirm or not his/her choice (Figure 7). They (the participant and the robot) could disagree or agree, and there was no “right answer”.

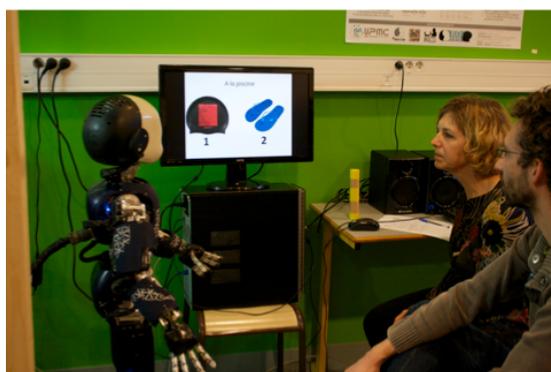
While in the functional task the evaluations on the perceptual characteristics of the stimuli were based on objective measures (even if difficult to discern in the ambiguous or equal case), here, evaluation was essentially based on a subjective and personal judgment. Hence, the strategy chosen by the experimenter for the robot answers was to always contradict the participant as done by Nass *et al.* (1996). Again, this required the operator to listen to the participant's answer and choose the appropriate answer each time, formulated as “the first one” or “the second one”.

---

<sup>7</sup> We submitted participants a limited number of questions (three) in the social tasks, for several reasons. One is to avoid the disengagement due to a long interaction with the robot, considering their frustration due to the fact that the robot is always contradicting them. Another reason was to avoid asking questions on matters where the participant could have had judgemental biases. Finally, as the functional task included three items (sound, color, weight), we chose to keep the same number of items for the social task in order to balance our experimental design.



**Figure 6a-b-c-** Social task: Q1: at school, which is the most important object: 1) the computer or 2) the notebook?; Q2: at the swimming pool, which is the most important object: 1) the bathing cap or 2) the flip-flops?; Q3: under the rain, which is the most important object: 1) the K-way or 2) the umbrella?



**Figure 7-** Social task: the experimenter interrogates the participant and the robot.

### 3.2.4. Procedure and Experimental Design

The present study was part of the Engagement During Human-Humanoid Interaction (EDHHI<sup>8</sup>) project. The human-robot interaction protocol applied in this project was validated by the local Ethical Valuation Council for Researches on Health Issues (CERES)<sup>9</sup>.

Volunteers who took part in the experiment were required to fill up the two questionnaires online at least two weeks before their visit to the Institute for Intelligent Systems and Robotics (ISIR). These two questionnaires were: (i) Negative Attitude towards Robots Scale (NARS; Nomura *et al.*, 2006) and (ii) Desire For Control scale (DFC; Burger *et al.*, 1979).

The day of the experiment, participants were welcomed by the researcher and informed about the overall procedure before signing an informed consent form granting us the use of all the recorded data for research purposes. Each participant was equipped with a Lavalier microphone to ensure a clear speech data collection.

<sup>8</sup> <http://www.smart-labex.fr/index.php?perma=EDHHI>

<sup>9</sup> IRB n.20135200001072

Before the experiment, the participants had to watch a short video presenting the iCub, its body parts and some of its basic movements.<sup>10</sup> The video did not provide any information about the experiments. It was instrumental to make sure that the participants had a uniform prior knowledge of the robot appearance (some participants may have seen the robot before on the media).

The 56 participants were all confronted to the two tasks. One group (G1) was asked to imagine a collaborative scenario (N = 21), while a second group (G2) was asked to imagine a competitive scenario (N= 21). Finally, the control group (G3) was instructed to imagine a neutral scenario (N= 14). The instructions related to the imagined scenarios varied according to the experimental conditions as follows. Instructions to elicit a collaborative (G1), competitive (G2) or neutral (G3) scenario are reported in Table 1. The five hypotheses were tested on the whole set of the participants.

---

<sup>10</sup> It is a dissemination video from IIT showing the iCub, available on Youtube (<http://youtu.be/ZcTwO2dpX8A>).

**Table 1** – The instructions used for the imagined scenario of the human-robot collaboration.

<b>Imagined scenario</b>	<b>English text</b>	<b>French text</b>
<i>Competitive</i>	<p><i>Imagine that in 2 years you will be working with a robot to build some objects: you will be in competition. The one that will have built the best object will win a prime.</i></p> <p><i>Imagine this scenario in detail for a minute.</i></p>	<p><i>Imaginez que dans deux ans vous travaillerez avec un robot pour construire des objets : vous serez en compétition. Celui qui aura construit le plus bel objet gagnera une prime.</i></p> <p><i>Imaginez ce scénario de façon détaillée, pendant une minute.</i></p>
<i>Neutral</i>	<p><i>Imagine that in 2 years you will work with a robot. You will have to build some nice objects.</i></p> <p><i>Imagine this scenario in detail for a minute.</i></p>	<p><i>Imaginez que dans deux ans vous travaillerez avec un robot. Vous devez construire de beaux objets.</i></p> <p><i>Imaginez ce scénario de façon détaillée, pendant une minute.</i></p>
<i>Collaborative</i>	<p><i>Imagine that in 2 years you will work with a robot to build some objects: you will make a duo with one of them and will have to collaborate to build some object. As a duo you can win a prime if the object is well built.</i></p> <p><i>Imagine this scenario in detail for a minute.</i></p>	<p><i>Imaginez que dans deux ans vous travaillerez avec un robot pour construire des objets : vous formez un binôme avec un d'entre eux et vous devez collaborer pour construire l'objet. En tant que binôme vous pouvez gagner une prime si l'objet est bien construit.</i></p> <p><i>Imaginez ce scénario de façon détaillée, pendant une minute.</i></p>

The participant was then introduced to the robot.

The experimenter did not present the experimental setup (e.g., show the location of the cameras) except showing the robot, and she/he did not provide any specific instruction to the participants about what to do or say and how to behave with the robot. Most importantly, the experimenter did not say anything about the way the robot was controlled: since the operator was hidden behind a wall, mixed with other students of the lab, the participant had no clue that someone else controlled the robot<sup>11</sup>.

The robot was standing on its fixed pole, gently waving the hands and looking upright. It was not speaking. Once the participant was standing and looking in front of the robot, s/he was free to do whatever s/he wanted: talk to the robot, touch it, and so on. The experimenter took seat on the right of the participant, in front of the robot and invited the participant to take the

<sup>11</sup> After the experiments, we asked the participants if they thought or had the impression that someone controlled the robot: all the participants thought that the robot was fully autonomous.

other seat. The experimenter then provided verbal instructions for the experiment, consisting of two tasks: a functional evaluation task, aimed at assessing the trust in the robot's *functional savvy*, and a social evaluation task, aimed at assessing the trust in the robot's *social savvy*. The participants executed the two tasks in a random order.

Upon task completion, participants were asked to rate, on a 7-points scale (1=very competitive; 7=very collaborative) whether they reminded the imagined scenario as competitive or collaborative to ensure that they actually imagined the proposed scenario. The interaction task lasted on average 30 minute per participant. The whole experiment took place in individual sessions in the experimental lab room of the iCub and it lasted on average 50 minutes for each participant. When the participant had finished, the experimenter thanked her/him and stored the collected data.

### 3.2.5 Data collection and analysis

The participants and the robot responses were recorded on an individual sheet by the experimenter, and could be additionally retrieved by the audio and video recordings.

Answers to the questions addressed during the functional and social tasks were used to create quantitative measures of participants trust in robot's functional and social savvy. The registered data consisted in the participants' conformation to, or disagreement with, the robot's answers.

A conformation score was calculated dividing the number of instances where the participant changed his/her answer to match the robot's answer (conformation) by the total number of instances where the robot's answer was in disagreement with the participant's first answer. Hence, acceptance score may take a value from 0 (disagreement) to 1 (conformation).

Responses to the two questionnaires were used to create quantitative measures of participants' desire for control and of attitudes towards social influence of robot. Scores to the NARS-S2 questionnaire were calculated in compliance with the method recommended by the authors (Nomura *et al.*, 2006). S2 score may range from 0 to 35. A high score indicates negative attitudes towards social influence of robots. Scores to the Desire for Control scale (Burger *et al.*, 1979) were calculated according to the authors' method. Score may range from 0 to 140. A high score indicates a strong desire for control. Over 56 participants, only 51 filled correctly the DFC questionnaire. The results based on the DFC scale are obtained by retaining

only these 51 participants.

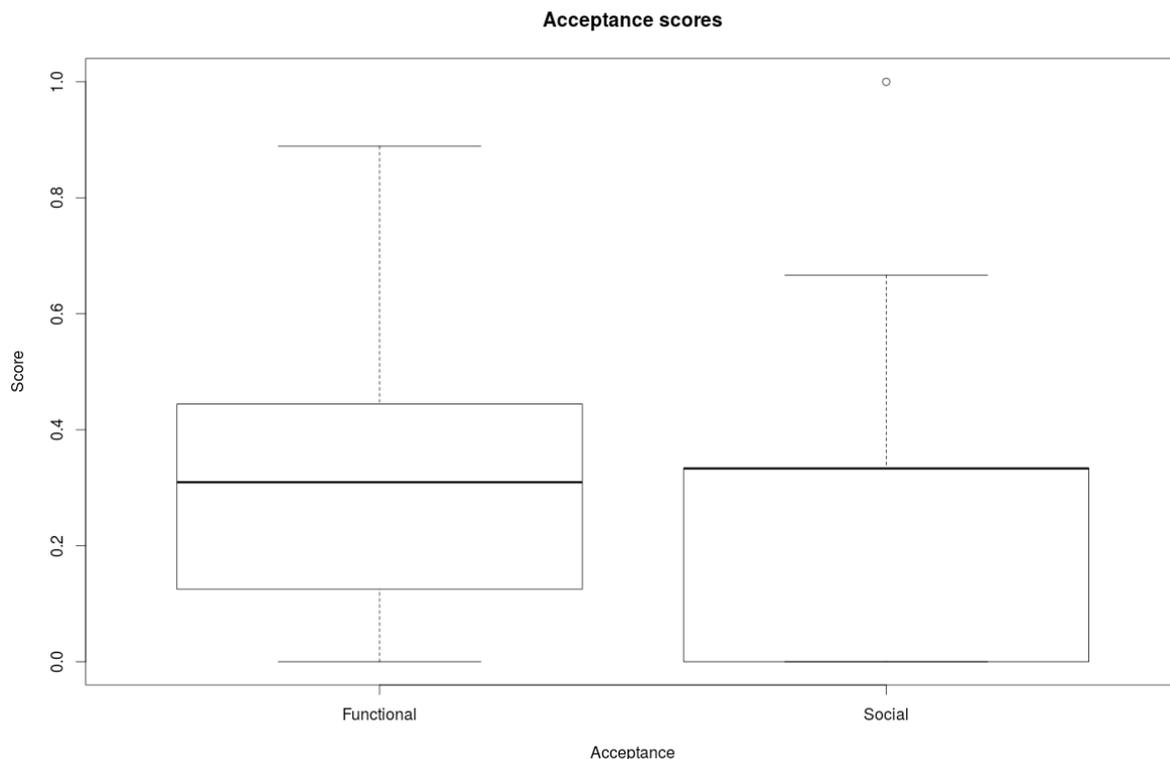
These three scores were used as dependent variables. As the functional and social conformation scores did not present normal score distributions, to carry statistical analysis we relied on non-parametric tests.

## 4. Results

### 4.1 Do participants conform more their answer to iCub answer in the functional task than in the social task? (H1)

The conformation score obtained on the functional and social questions ranked from 0 (never conform) to 1 (always conform). In order to calculate to what extent participants trust robots during functional and social tasks, we have considered a score threshold of 0.5, which is the middle of the score scale, ranging from 0 to 1. We have thus esteemed that a conformation score higher or equal to 0.5 in each of the two tasks reveals a participant's trust in the robot's savvy.

Descriptive analysis performed on the 840 [=56\*(12+3)] participants' answers reveal that the average conformation score for the functional task is 0.315 ( $\sigma=0.201$ ), while the average conformation score for the social task is 0.199 ( $\sigma=0.221$ ). In both functional and social tasks, the mean score is lower than 0.5 indicating that overall participants do not conform so easily to the robot answers. However, the Wilcoxon test for paired samples performed lets emerge a significant difference ( $V=981$ ;  $p<.001$ ) between these two scores. Participants tend to conform more to the robot's functional than social answers (Figure 8). Hence, our first hypothesis is confirmed: participants trust more in the iCub's functional savvy than in its social savvy.



**Figure 8** - Conformation score in the functional and social tasks for the 56 participants.

## 4.2 Do participants who conform to iCub in the social task also conform in the functional task? (H2)

Among the 56 participants, 13 have obtained a high functional conformation score ( $\geq 0.5$ ) while only 3 participants showed a social conformation score higher or equal to 0.5<sup>12</sup>. Table 2 outlines the conformation score of these three participants for both functional and social tasks.

**Table 2:** Conformation score in functional and social tasks of the 3 participants who showed trust in social savvy

Participant	Conformation score	
	<i>Functional task</i>	<i>Social task</i>
#57	0.333	0.667
#76	0.333	0.667
#101	0.5	1.0

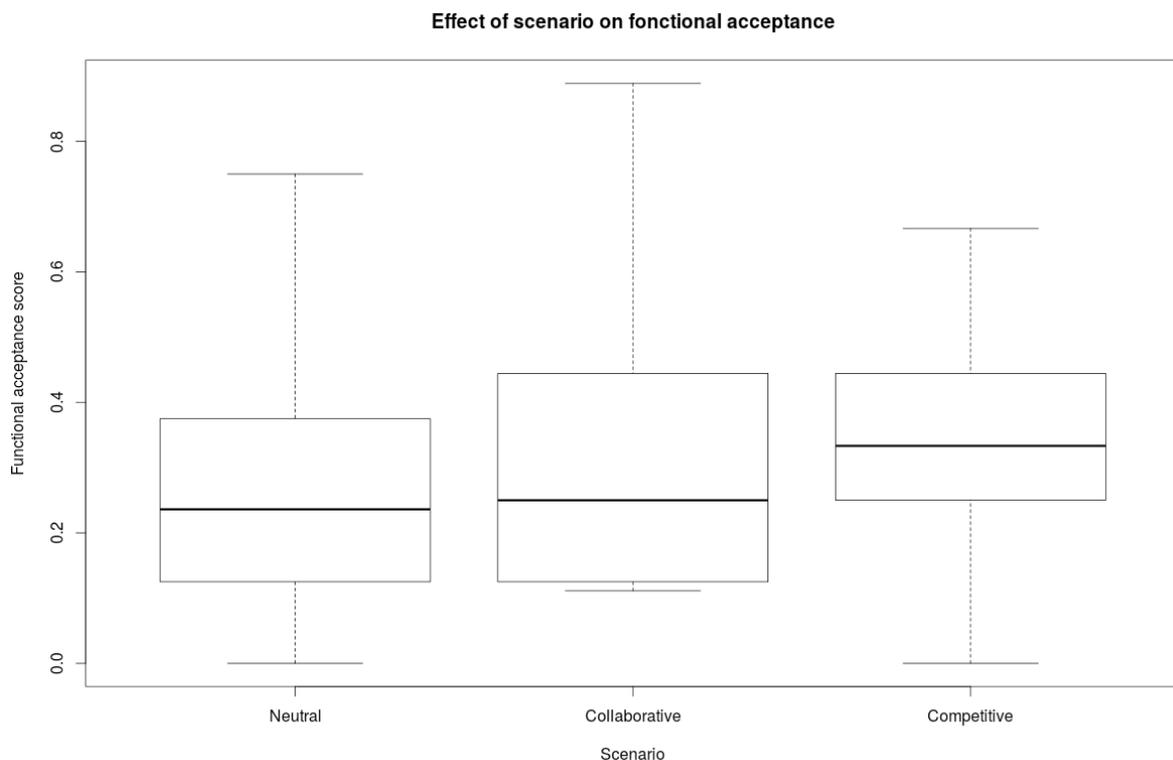
<sup>12</sup> Our data show that the majority of our participants (92.8%) present a low conformation score in the social task, revealing the following distribution: (always disagree) 0 (48.2%), 0.33 (44.6%), 0.5 (1.8%), 0.67 (3.6%) and 1(1.8%) (always conform).

Such a low number of participants showing a high conformation score in the social task does not allow to carry statistical tests. However, we may observe that two of these three participants have also obtained a lower score for functional task and one a score equal to 0.5, which is the threshold of low/high score. These results tend not to confirm our second hypothesis, *i.e.*, that participants do not trust social savvy uniquely, so that those who conform to the robot in the social task also would have conformed in the functional task.

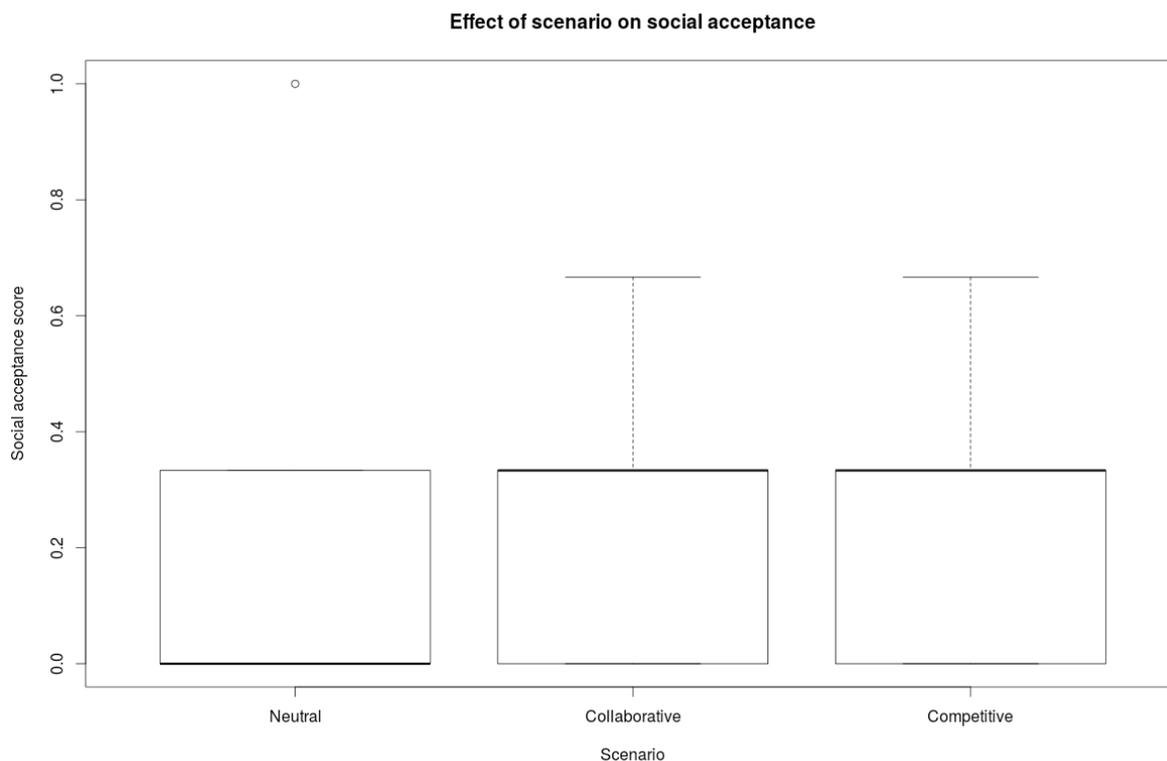
### 4.3 Does the imagined HRI scenario influence trust in iCub? (H3)

To assess whether the two scenarios had an impact on participants conformation we have performed a non-parametric ANOVA (Kruskal-Wallis test). The main between-subject factor is the scenario condition (3 levels: collaborative, competitive and neutral). The dependent variables are the functional and social conformation scores.

Results do not show any effect of the imagined HRI scenario on participants conformation score in the functional task ( $\text{Chi}^2=1.69$ ; *n.s.*) – see Fig.9 – nor in the social task ( $\text{Chi}^2=1.63$ ; *n.s.*) – see Fig. 10. Thus, our fourth hypothesis was not confirmed: imagining a collaborative interaction with robot does not seem to increase the trust in the robot.



**Figure 9** - Effect of the imagined HRI scenario on the participants' conformation score in the functional task.



**Figure 10** - Effect of the imagined HRI scenario on the participants' conformation score in the social task.

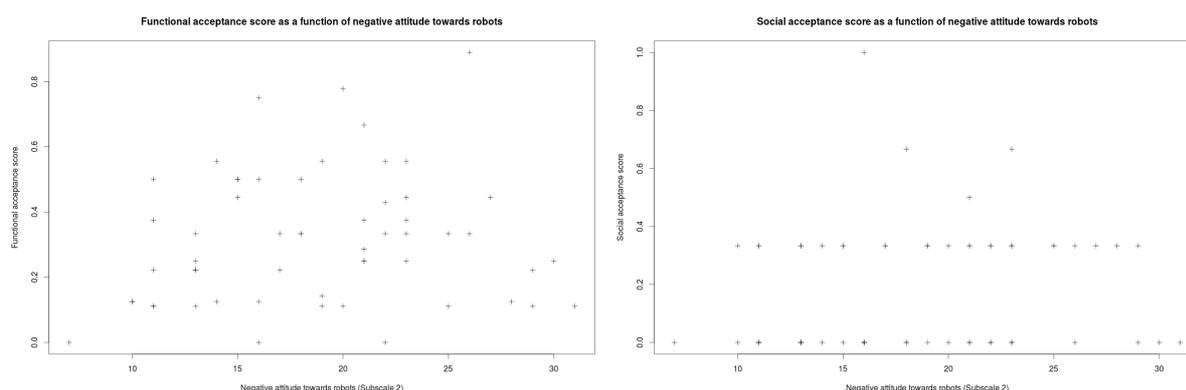
After the task, we checked whether participants remembered the type of scenario described by the experimenter (collaborative, competitive, and neutral scenario). To do so, we asked the participants to rate, on a 7-points scale (1=very competitive; 7=very collaborative), whether they reminded the kind of scenario they have imagined. We carried out a non-parametric ANOVA with this score as dependent variable and the imagined scenario as factor. Results show a significant difference among the three imagined scenario ( $M_{\text{compet}}=3.57$ ;  $M_{\text{neutral}}=4.76$ ;  $M_{\text{collabe}}=5.21$ ;  $\text{Chi}^2=13.83$ ;  $p<.001$ ). Moreover the Wilcoxon post-hoc tests allow us to observe that there is a significant difference between competitive and neutral scenario ( $W=325.5$ ;  $p<.001$ ), and between competitive and collaborative scenario ( $W=245.5$ ;  $p<.001$ ). While no significant difference has been registered between collaborative and neutral scenario ( $W=177$ ; *n.s.*).

These results show that the participants in the competitive group tend to remind the scenario as less competitive than actually described in the instructions given by the experimenter. Therefore, our results showing no effect of the scenario on functional (Fig. 9) and social (Fig. 10) conformation should be taken carefully.

#### 4.4. Is there a correlation between negative attitudes to robot's social influence and the trust in the robot's social savvy? (H4)

Descriptive analysis of the NARS-S2 (subscale 2 of NARS) scores shows that they follow a normal distribution (Shapiro test:  $W=.977$ ;  $n.s.$ ):  $M=18.80$ ;  $SD=5.83$ . In order to identify a potential correlation between the conformation score and those of the NARS-S2, we performed a Spearman non-parametric test of correlation (Fig.11). Results do not show any significant correlation between the conformation score in the functional task and the score assessing the negative attitude towards the social influence of robots (NARS-S2) ( $\rho=.127$ ;  $n.s.$ ) nor between the conformation score in the social task and the NARS-S2 attitude score ( $\rho=.127$ ;  $n.s.$ ). No difference on this correlation was observed in each of the 3 imagined scenarios.

These results confirm the hypothesis that a negative attitude towards social influence of robots negatively correlates with the trust in the robot's social savvy expressed by the conformation score.



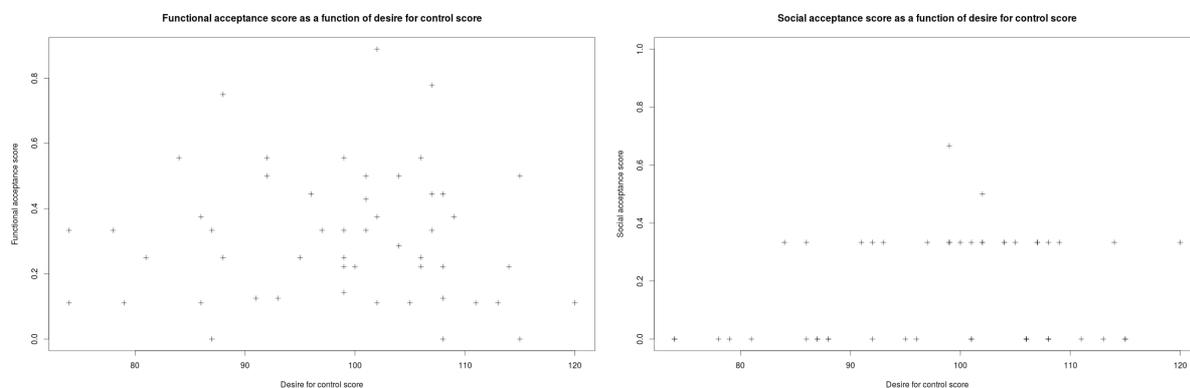
**Figure 11** - Correlation matrix between the conformation score of the functional and social task and the scores of the NARS-S2.

#### 4.5 Is there a correlation between the human desire for control and the trust in the robot's functional and social savvy? (H5)

Descriptive analysis of the DFC scores shows that they follow a normal distribution (Shapiro test:  $W=.964$ ;  $n.s.$ ):  $M=98.7$ ;  $SD=11.1$ . To identify a potential correlation between the conformation scores and the “desire for control” (DFC) scores, we performed a non-parametric correlation Spearman test (Fig. 12). Results do not show any correlation between the conformation scores in the functional task and the DFC score ( $\rho=-.086$ ;  $n.s.$ ), nor between the conformation scores in the social task and the DFC score ( $\rho=.137$ ;  $n.s.$ ). No difference in

this correlation was observed on each of the 3 imagined scenarios.

These results do not confirm the hypothesis that desire for control negatively correlates with the trust in the robot's functional and social savvy expressed by the conformation score.



**Figure 12** - Correlation matrix between the conformation score for functional and social task and those of the DFC.

Finally we observe a negative correlation between the DFC and the NARS-S2 ( $r=-.443$ ;  $p<.005$ ). That is the more participants have a high desire for control, the less they show negative attitudes towards social influence of the robot.

## 5. Discussion

Based on the studies showing that the user's trust is a fundamental ingredient of HRI and can thus be employed as an indicator of robot acceptance (Schaefer, 2013; Ososky *et al.*, 2013; Hancock *et al.*, 2011; Parasuraman *et al.*, 1997), the present study aims at gaining insights into the human-robot trust dynamics that could affect the acceptance of robots in daily situations.

In particular, we were interested at investigating trust in functional savvy (Fridin *et al.*, 2014; Heerink, 2010; Shaw-Garlock, 2009) as an indicator of users acceptance of the robot as functional agent, and trust in social savvy (de Graaf *et al.*, 2015; Weiss *et al.*, 2009; Young *et al.*, 2009) as an indicator of acceptance of the robot as a social agent.

While several empirical studies prove that people do trust computers social savvy during human-computer collaborative interaction (Nass *et al.*, 1996; Nass *et al.*, 1995), it is still not clear if this holds true for human-robot and particularly for human-humanoid interaction. Indeed, while users consider functional savvy as indispensable for accepting the robot (Fridin *et al.*, 2014; Beer *et al.*, 2011; Heerink *et al.*, 2009; Weiss *et al.*, 2009; Kaplan, 2005), social

savvy is rather desirable (Lohse, 2010; Dauthenhan, 2007) – but sometimes deceptive (Coeckelbergh, 2012; Duffy, 1993).

Thus, the first purpose of our study was to find out whether users trust robots both when required to take decisions on functional issues and when required to do it on social issues. The Media Equation Theory paradigm (Nass *et al.*, 2000) provides a sound basis to investigate trust in technology by registering users' conformation or disagreement with computers. Based on this experimental paradigm and on above presented research on robots acceptance, we hypothesized that trust in functional savvy should be more common than trust in social savvy and predicted that participants would have conformed to iCub answers more in functional than in social decision-making tasks.

Results show that in general participants do not easily conform to iCub answers. However, those who conformed mostly did it in the functional task. Hence, our first hypothesis was partially confirmed: participants tended to see robots as untrustworthy, or at least not enough trustworthy to take better decisions than humans, except when these decisions concern functional issues which require high-precision technical skills to discriminate a specific perceptual characteristic of the stimuli. In this case the participants relied on the robot's functional savvy and accepted more often that iCub's answer determines their final decision. The analysis of verbal registrations of the participants' answers confirms this reliance: several participants reported that they preferred to rely on iCub's perception than on their own perception. Furthermore, these registrations led us to remark that participants who did not conform in the functional task motivated their lack of conformation referring to their expertise in a specific domain. For example, one participant said that she was a musician, and so she felt more confident on her own ability to distinguish sounds than on the one of the robot.

The differences observed among the participants' trust in the robot's functional and social savvy could then be explained by the fact that, since functional savvy is based on a objective knowledge and on high technical skills which participants tacitly ascribe to robots as a core knowledge (Kaplan, 2005; Beer *et al.*, 2009; Heerink *et al.*, 2009), the robot is more easily accepted as a tool for functional duties. Whereas, because social savvy is based on a common-sense knowledge that is usually acquired through repeated experience in social contexts (Guimelli, 1999), it can be implausible for users to assume that such kind of savvy is instilled in the robot like a built-in knowledge or setting, or to believe that robot has a life history behind it, made of social interactions and situations. Consequently the robot is hardly accepted as a trustworthy partner in social tasks, at least at the current stage of the

introduction of robots in social contexts. To this concern, commentaries of participants retrieved in the verbal registration witness that some participants raised his/her concern about the credibility of the robot when answering social questions. For example, participants remarked that the robot had never gone to the swimming pool, so it could not know which was the most appropriate item to use in that context. Similarly, several participants remarked that the robot had never been under the rain so it could not know which was the most appropriate item in that context neither.

Another possible explanation is that while the validity of the decisions taken to achieve functional tasks can be corroborated by the physical reality, (*i.e.*, the impression of « good or wrong estimation » can be confirmed or not by scientific evidence - see Kahneman, 2003 for a review) the validity of decisions taken to achieve social tasks (*i.e.*, the judgment of « good or wrong doing ») can be validated only inter-subjectively (Nichols, 2004). Therefore, while in functional tasks participants decision-making is achieved with the awareness that they are not the “final judge” of the perceived reality, so that they can leave the decision to the robot, social tasks are more likely to arouse a persuasive behavior, which leads participant to give more importance to their personal judgments and to defend them inter-subjectively (*i.e.*, in front of iCub) as the best decision in a given context. In fact, while performing the social task, several participants commented that they wish to express their own opinion even when knowing that robot’s answer was the correct one.

Moreover, looking at this result in the light of those obtained by Nass *et al.* (2000), we can point out that, despite human-like robots are increasingly conceived and perceived as social technologies (Barnteck *et al.*, 2007), we did not registered a comparable or more significant level of social trust with relation to robots than the one registered by Media Equation Theory studies with relation to computers. These results suggest that, currently, users trust more the computer social savvy than the robot’s social savvy. If at a first glance this may seem counterintuitive, we should consider that while we have today little familiarity with robots, our usage of computers is frequent and fluent. It is then possible that we trust more computers because we are already acquainted to them, and thus we tend to believe that no additional effort is needed to use them for social tasks. Furthermore, while a computer is characterized by powerful computation functionalities that can support decisions (Davis, Bagozzi & Warshaw, 1989), the embodied and behavioral nature of robots potentially enable them to transform computations and decision into actions. It is thus reasonable to think that the robots’ social savvy may be seen as having more profound and dangerous consequences than the

computers' social savvy, and this can cause a distrust bias as a manifestation of the anxiety towards the robots during social tasks. However, contrary to other experimental paradigms where participants are asked to take decisions that imply concrete consequences (see Khaneman, 2007) or that confront them to "moral dilemma" (e.g., Malle, Scheutz & Voiklis 2015) in our study the decisions taken by participants do not have any consequential effect on real actions nor on morality. Therefore to validate this interpretation of the results, in future research we shall introduce decision-making task whose objective is not only to provide a good answer but also to care for the consequences of this answer.

Interestingly, not only trust in robot is different from trust in computers, but also from trust in humans (Billings, Schaefer, Llorens *et al.*, 2012; Mayer *et al.*, 1995; Rotter, 1971). It has been proven that, although users are rather inclined to project social life on robots, *i.e.*, they tend to interpret robot behavior as they do with humans or animals by attributing animacy and social skills to them (Severson & Carlson, 2010), they do not attribute free will and intentionality to robots, *i.e.*, they do not believe that robots are capable of voluntary choice and social judgment (Monroe, Dillon & Malle, 2014; Montague, 2008). It is thus possible that in our study participants did not conform to the iCub because they did not attribute it neither the capability of voluntary choice nor the common-sense knowledge or social judgment that is underpinned by this capability. The resulting portrait of the robot is thus the one of a social but not intentional agent - and maybe this incongruous status is what makes the uniqueness of the robot.

In this sense, trust in robots may present specificities with relation to trust in computers and humans so that these specificities can be further investigated in future research by comparative experimental conditions involving human-computer, human-robot interactions and human-human interactions. In particular, given the child-like aspect of iCub, it would be interesting to compare interactions between two children *vs.* interaction between a child and iCub.

To continue on our findings, following the proposal of Young *et al.* (2009) that trust in functional savvy is a prerequisite for trust in social savvy, our second hypothesis was that users would not trust the social savvy uniquely. Consequently, we predicted that participants conforming to the iCub's answers in the social task would also conform in the functional task. Contrary to our hypothesis and to the current literature (Young *et al.*, 2009), results show that the very limited number of participants conforming in social task conformed less in the functional task. This result indicates that the few participants who believed that social savvy is

susceptible to be an intrinsically reliable ability of the robot, do not believe the same for functional savvy. This minority does not base trust in social savvy on trust in functional savvy, and on the contrary seems to assume that a social robot that has poor technical skills and little objective knowledge would not be necessary an untrustworthy social robot.

Globally, this first set of results leads us to conclude that, over the limited number of participants who trusted in the robot, a restrained number considered that trust in functional savvy excludes trust in social savvy, and an even more restrained number of users who considered that trust in social savvy exclude trust in functional savvy. This result could be symptomatic of a dichotomist view of the robot by the participants: either the robot is a machine that can assist the human in functional tasks, or it is a companion that can interact at a social level.

With regards to the desire for control, the negative attitude towards the robots' social influence, and the simulated interaction scenario with robot, the results only partially confirmed our hypothesis as well as the discussed literature.

More in detail, results concerning the influence of the desire for control on the participants conformation indicate that, whether participants showed a high or low desire for control, they mainly tended not to conform to iCub. This confirms previous studies where users generally preferred to be in control of the robot (Koay *et al.*, 2014; Kamide *et al.*, 2013; Okita *et al.*, 2012; Syrdal *et al.*, 2007; Gilles *et al.*, 2004; Marble *et al.*, 2004). With regards to the thirteen participants conforming to iCub in the functional task, and to the three participants conforming to iCub in the social task, we might ascribe this behavior to the fact that the young age of iCub, whose appearance was that one of a 4 years old child, aroused a tolerant (if not rather positive) behavior, so that these participants did not wish to contradict iCub just like sometimes parent or teacher find it difficult to contradict a child. For example, one participant reported that she wished not to contradict the robot because it was cute, and she wanted to please it. When asked, she said the robot reminded her of a small child.

Concerning results on the influence of the attitude towards robots, results show that the participants' trusting behaviors were independent from negative attitude towards the social influence of the robot: whether they fear to be influenced by the robot or not, participants did not conform their decisions to those of the robot, except for the minority who conformed during functional tasks and for the even more restrained minority who conformed during social tasks. This could mean that distrust is itself a deeply rooted and widely diffused attitude

towards robot, more than specific fears about robots like the fear of being influenced by them. Hence, even when these fears will be overcome, if they will, still psychologists and engineers will have to deal with a generalized and substantial level of distrust amongst users.

Further analysis have also showed that the more participants have an elevated desire for control (score at the DFC), the less they fear to be influenced by robots (score at the NARS-S2). These participants seem thus not to be concerned by the possibility that a robot could determine their decisions as they probably assume that they will generally be in control of their functional and social tasks and they would not need to rely on robots. This result seems to indicate that participants with strong desire for control interpret trust mainly as reliance, and that the idea of being in the condition to rely on a robot makes them unwilling to use the robot. Correlation between desire of control and negative attitudes towards robot could thus turnout to be revealing of an intention to use the robot. Intention to use the robot pre-exists any eventual trust behavior, and thus supplies an early evidence of a tendency to resistance, rather than to acceptance. Prospective developments of the present study should thus identify the participants' levels of intention of use as a preliminary predictor of trust behavior. Moreover since this study mainly focuses on robots' functional and social acceptance, it shall be useful to additionally assess participants regular use of technology, by asking them if they normally use social technologies (*e.g.*, social networks, videogames, etc.), and whether they use functional technologies for social purposes (*e.g.*, using an office computer for chatting, watching a movie, etc.) as well as social technologies for functional use (*e.g.*, using social networks to retrieve useful information, selling objects, looking for a job, etc.). This information might help to predict trust behaviors in future research.

Finally, results of the influence of the imagined scenario on trust behavior seem to infirm our hypothesis that imagining a collaborative interaction scenario would have determined a higher level of trust, and while this in contrast with the findings in the study of Kuchenbrandt *et al.* (2012) witnessing that robots acceptance increases when users imagine a collaborative scenario, this is in line with the more recent study of Wullenkord *et al.* (2014), where participants who had imagined contact with a robot did not report more positive attitudes towards robots. However, in the specific case of our study, the weak influence of the imagined scenario on participants' level of trust can be explained by the fact that participants had difficulties in imagining the proposed interaction scenario. Apparently, participants who had been asked to imagine a competitive scenario could hardly recall that they imagined it. This can be due to the fact that while in the imagined scenario participants thought about an

abstract non-specified robot, in the real interaction they were confronted with a specific robot that has a child-like appearance. Again, the appearance of iCub could have diminished the strength of the competitive scenario and enhanced the strength of collaborative scenario in their memory. Therefore, in future research we shall vary the physical appearance of a robot to validate our hypothesis on the influence of imagined scenario on users trust.

## 6. Conclusions

This study shows that robot acceptance is a complex dynamics characterized by a prevailing distrust in robots, and where the limited trust behaviors that can be observed amongst users are significantly correlated to the nature of the task at hand. In particular, robots seem to be more easily accepted in functional than in social tasks. This is witnessed by the evidence that when confronted with tasks requiring decisions about functional issues, users trust robots more than they do when confronted with tasks requiring decision on social issues. Moreover, the minority of users who trust robots on social issues show a significant distrust in robots on functional issues.

These results do not allow to understand whether the observed distrust and trust behaviors depend on the fact that, despite today robot's vocation of being at the same time a functional and social technology, users still consider robots as "socially ignorant" (Young *et al.* 2009) or rather on the fact that trust in functional savvy and trust in social savvy are mutually exclusive because they are based upon different kinds of knowledge and skills (scientific objective knowledge and technical skills *vs.* subjective common-sense knowledge and adaptive skills based on social norms). However, the general distrust and the different behavior of participants in the two tasks suggest that trust in robots cannot be assimilated to mechanisms that are typical of trust in computers neither to those that are typical of trust in humans.

Furthermore, the three observed HRI factors (desire for control, negative attitudes towards social influence of robots, and imagined interaction scenario) seem not to have influence on trust behaviors in terms of conformation of users' decisions to robot decisions. Nonetheless, we registered a significant correlation between desire for control and negative attitudes towards robot. This correlation indicates that the more participants wish to have control on their life situations, the less they fear of being influenced by a robot, and this suggests that users with control-seeker profiles do not wish to use robots as far as they would have to rely on them. Consequently, we argued that correlation between desire for control and negative attitudes towards robots could help in quantifying the individual's intention of use.

Furthermore, we proposed that the intention of use can be an interesting predictor of trust for participants whose strong desire for control implies that they reduce the complexity of trust to one of its component, which is reliance (*e.g.*, dependence from robots), and tend thus to resistance to robots rather than to accept them. Additionally, we have proposed that users' habits with concerns to functional and social technologies can also be investigated as predictors of trust behavior.

Finally, from a theoretical and methodological point of view, our study is an attempt to establish a conceptual bond among pivotal issues that, though being recurrently used in the HRI communities such as “social and functional acceptance”, “robots savvy” and “trust”. These issues, though being recurrent used in the HRI communities still suffer of some fuzziness since they have be inherited from other domains, namely ergonomics and social psychology, and need to be adapted to social robotics in order to get a more profound understanding of users acceptance of robots. To understand what kind of robots acceptance we are talking about (functional vs. social) is now an urgent issue because while people could “blindly” trust robots due to their increased automation, artificial intelligence and related functional abilities, they could eventually ignore signs of malfunction (see for instance our current behavior with GPS), or they could, on the opposite, doubt of the robot's social competence with respect to more sensitive issues such as for instance those related to healthcare situations (*i.e.*, taking treatments when suggested).

To this concern, today people seem to be confronted with a choice between two possible user profiles, namely a strong technology-mediated profile – characterized by the use of robots or other technologies as intermediaries in their personal and professional contexts, tools, and practices – and a strong non-technology-mediated profile – characterized by a quest for a return to “direct relations with things”. In between, societal initiatives like that one of the Fablabs (fabrication laboratories, where users can for example build a robot by themselves – *e.g.*, Walter-Herrmann & Büching, 2014) propose an alternative user profile and a more aware relationship with technology, that reflect both an ideological posture and an economical response with respect to the technology overproduction and to the subsequent irresponsible consuming, in order to prevent the unfair consequences of a huge technological gap between people who buy – and thus rely on - robots and people who don't.

In this sense, while so far “dependability” has been considered as a feature of the robot (Hancock *et al.*, 2011), it becomes more and more crucial to investigate dependability as a

user feature. Moreover, as Serva and colleagues have pointed out (Serva, Fuller & Mayer 2005) despite users' trust in an artificial partner is susceptible to develop through the process of being trusted by that partner, a phenomenon termed 'reciprocal trust', such phenomenon stays still underexplored in the literature. Finally, further studies on acceptance and trust may help to prevent an eventual radicalization of users' behavior, either towards dependence from a technology or towards a refusal of it.

## References

- Abric, J. C. (1994a). *Représentations sociales et pratiques*. Paris: PUF.
- Abric, J. C. (1994b). L'organisation interne des représentations sociales: système central et système périphérique. In C., Guimelli. *Structure et transformation des représentations sociales*.
- Alavi, M., & Joachimsthaler, E. A. (1992). Revisiting DSS Implementation Research: A Meta-Analysis of the Literature and Suggestions for Researchers. *MIS Quarterly*, 16 (1), 95–116.
- Alport, G.W. (1935). Attitudes. In C. Murchison (Ed.), *Handbook of social psychology* (pp. 798–844). Worcester, MA: Clark University Press.
- Baddoura, R., & Venture, G. (2013). Social vs. Useful HRI: Experiencing the Familiar, Perceiving the Robot as a Sociable Partner and Responding to Its Actions. *International Journal of Social Robotics*, 5, 529–547.
- Bartneck, C., Suzuki, T., Kanda, T., & Nomura, T. (2007). The influence of people's cultures and prior experiences with Aibo on their attitude towards robots. *AI & Society*, 21, 217–230.
- Beer, J. M., Prakash, A., Mitzner, T. L., & Rogers, W. A. (2011). Understanding Robot Acceptance. *Technical Report HFA-TR-1103*. Atlanta, GA: Georgia Institute of Technology School of Psychology – Human Factors and Aging Laboratory. Retrieved from: <http://hdl.handle.net/1853/39672>
- BenMessaoud, C., Kharrazi, H., & MacDorman, K.F. (2011). Facilitators and Barriers to Adopting Robotic-Assisted Surgery: Contextualizing the Unified Theory of Acceptance and Use of Technology. *PLoS One*, 6(1).
- Billings, D. R., Schaefer, K. E., Chen, J. Y., & Hancock, P. A. (2012). Human-robot interaction: developing trust in robots. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction* (pp. 109-110). ACM.
- Billings, D. R., Schaefer, K. E., Llorens, N., & Hancock, P. A. (2012). *What is Trust? Defining the construct across domains*. Poster presented at the American Psychological Association Conference, Division 21. Orlando, FL.
- Biros, D., Daly, M., & Gunsch, G. (2004). The influence of task load and automation trust on deception detection. *Group Decision and Negotiation*, 13, 173–189.
- Burger, J. M., & Cooper, H. M. (1979). The desirability of Control. *Motivation and Emotion*, 3, (4), 381–393.
- Burke, C. S., Sims, D. E., Lazzara, E. H., & Salas, E. (2007). Trust in Leadership: A multi-level review and integration. *The Leadership Quarterly*, 18, 606–632.
- Cameron, D. Aitken, J.M., Collins, E.C., Boorman, L., Chua, A., Fernando, S., McAree, O., Martinez-Hernandez, U., & Law, J. (2015). *Framing Factors: The Importance of Context and the Individual in Understanding Trust in Human-Robot Interaction*. Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2015), Workshop on Designing and Evaluating Social Robots for Public Settings.
- Carnap, R. (1942). Introduction to semantics. Cambridge-Massachusetts: Harvard University Press.
- Charalambous, G., Fletcher, S., & Webb, P. (2015). The Development of a Scale to Evaluate Trust in Industrial Human-robot Collaboration. *International Journal of Social Robotics*, 1-17.
- Coeckelbergh, M. (2012). Can We Trust Robots? *Ethics and Information Technology*, 14 (1), 53–60.
- Crisp, R. J., & Turner, R. N. (2012). Imagined intergroup contact. In G. Hodson & M. Hewstone (Eds.), *Advances in Intergroup Contact* (pp. 135–151). Hove: Psychology Press. Taylor & Francis.

- Dahl, T. S., & Boulos, M. N. K. (2013). Robots in Health and Social Care: A Complementary Technology to Home Care and Telehealthcare? *Robotics*, 3(1), 1-21.
- Davis, F., Bagozzi, R., & Warshaw, R. (1989). User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. *Management Science*, 35, 982–1003.
- de Graaf, M.M.A., & Allouch, S. B. (2013) Exploring influencing variables for the acceptance of social robots. *Robotics and Autonomous Systems*, 61, 1476–1486.
- Desai, M., Stubbs, K., Steinfeld, A. & Yanco, H. (2009). Creating Trustworthy Robots: Lessons and inspirations from automated systems. *Proceedings of the AISB Convention: New Frontiers in Human-Robot Interaction*, Edinburgh, Scotland. Retrieved from [https://www.ri.cmu.edu/pub\\_files/2009/4/Desai\\_paper.pdf](https://www.ri.cmu.edu/pub_files/2009/4/Desai_paper.pdf)
- DeSteno D., Breazeal, C., Frank, R. H., Pizarro, D., Baumann, J., Dickens, L., & Lee, J. J. (2012). Detecting the trustworthiness of novel partners in economic exchange. *Psychological Science*, 23, 1549–1556.
- Destephe, M., Brandao, M., Kishi, T., Zecca, M., Hashimoto, K., & Takanishi, A. (2015). Walking in the uncanny valley: importance of the attractiveness on the acceptance of a robot as a working partner. *Frontiers in Psychology*, 6, 204.
- Dillon, A. (2001). User Acceptance of Information Technology. In W. Karwowski (Ed.) *Encyclopedia of Human Factors and Ergonomics*. London: Taylor and Francis.
- Dinet, J., & Vivian, R. (2014). Exploratory investigation of attitudes towards assistive robots for future users. *Le travail humain*, 77, 105–125.
- Duffy, B. (2003). Anthropomorphism and the social robot. *Robotics and Autonomous Systems*, 42, 177–190.
- Eder, K., Harper, C., & Leonards, U. (2014). Towards the safety of human-in-the-loop robotics: Challenges and opportunities for safety assurance of robotic co-workers'. *Proceedings of IEEE International Workshop on Robot and Human Interactive Communication 2014 (IEEE RO-MAN 2014)* (pp. 660–665). Institute of Electrical and Electronics Engineers, Inc. (IEEE).
- Evans, A. & Revelle, W. (2008). Survey and behavioral measurements of interpersonal trust. *Journal of Research in Personality*, 42(6), 1585–1593.
- Fischer, K. (2011). Interpersonal variation in understanding robots as social actors. In *Proceedings of International Conference on Human-Robot Interaction* (pp. 53–60). Lausanne, Switzerland.
- Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and Autonomous Systems*, 42, 143–166.
- Freedy, A., de Visser, E., Weltman, G., Coeyman, N. (2007). Measurement of trust in human-robot collaboration. In: *International Symposium on Collaborative Technologies and Systems (CTS 2007)*, pp. 106–114
- Fridin, M., & Belokopytov, M. (2014) Acceptance of socially assistive humanoid robot by preschool and elementary school teachers. *Computers in Human Behavior*, 33, 23–31.
- Fumagalli, M., Ivaldi, S., Randazzo, M., Natale, L., Metta, G., Sandini, G., & Nori, F. (2012). Force feedback exploiting tactile and proximal force/torque sensing. Theory and implementation on the humanoid robot iCub. *Autonomous Robots*, 33 (4), 381–398.
- Gaudiello, I., Lefort S., & Zibetti, E. (2015). The place of robots in common-sense ontology and their prospected educative roles: how firm our representations are? *Computers in Human Behavior*, 50, 259–273.
- Gaudiello, I. (2015). *Learning robotics, with robotics, by robotics. A study on three educational paradigms of Educational Robotics, under the issues of robot representation, robot acceptance, and robot impact on learning*. PhD Dissertation, Université de Paris 8, Paris, France.
- Groom, V., & Nass, C. (2007). Can robots be teammates? Benchmarks in human-robot teams. *Interaction Studies*, 8 (3), 483–500.
- Guimelli, C. (1999). *La pensée sociale*. Paris: Presses Universitaires de France.
- Hamill, L., & Harper, R. (2006). Talking intelligence: a historical and conceptual exploration of speech-based human-machine interaction in smart homes. In *Proceedings of the International Symposium on Intelligent Environments (ISIE 2006)* (pp. 121–128). Cambridge, UK. Cambridge: MSR Press.
- Hancock, P.A., Billings, D.R., Schaefer, K.E., & Chen, J.Y., de Visser E.J., & Parasuraman R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors*, 53 (5), 517–27.
- Heerink, M. (2010). *Assessing acceptance of assistive social robots by aging adults*. PhD Dissertation, University of Amsterdam.
- Heerink, M., Kröse, B. J. A., Wielinga, B. J., & Evers, V. (2010). Assessing acceptance of assistive social agent technology by older adults: The Almere model. *International Journal of Social Robotics*, 2, 361–375.
- Heerink, M., Krose, B., Wielinga, B., & Evers, V. (2009). Influence of social presence on acceptance of an assistive social robot and screen agent by elderly users. *Advanced Robotics*, 23, 1909–1923
- Hinds, P. J., Roberts, T. L., & Jones, H. (2004). Whose Job Is It Anyway? A Study of Human–Robot Interaction in a Collaborative Task. *Human-Computer Interaction*, 19, 151–181.
- Hoffman, R. R., Lee, J. D., Woods, D. D., Shadbolt, N., Miller, J., & Bradshaw, J. M. (2009). The Dynamics of Trust in Cyberdomains. In R. R. Hoffman, J. M. Bradshaw, & K. M. Ford (Eds.), *Human-Centered*

- Computing (pp. 5–11). IEEE Computer Society.
- Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A Meta-Analysis on the Correlation Between the Implicit Association Test and Explicit Self-Report Measures. *Personality and Social Psychology Bulletin*, 1369–1385.
- Jack, A. I., Dawson, A. J., Begany, K. L., Leckie, R. L., Barry, K. P., Ciccia, A. H., & Snyder, A. Z. (2013). fMRI reveals reciprocal inhibition between social and physical cognitive domains. *NeuroImage*, 66, 385–401.
- Joose, M., Sardar, A., Lohse M., & Evers, V. (2013). BEHAVE-II. A Revised Set of Measures to Assess Users' Attitudinal and Behavioral Responses to a Robot's Social Behaviors. *International Journal of Social Robotics*, 5 (3), 379–388.
- Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, 58, 697–720.
- Kamide, H., Mae, Y., Takubo, T. Ohara, K., & Arai, T. (2014). Direct Comparison of Psychological evaluation between Virtual and Real Humanoids; Personal space and subjective impressions. *International Journal of Human-Computer Studies*, 72 (5), 451–459.
- Kaniarasu, P., Steinfeld, A., Desai, M., & Yanco, H. A. (2012). Potential measures for detecting trust changes. *Proceedings of the 7th Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI 2012)* (pp. 241–242), Boston, Massachusetts.
- Kaplan, F. (2004). Who is afraid of the humanoid? Investigating cultural differences in the acceptance of robots. *International Journal of Humanoid Robotics*, 1 (3), 1–16.
- Kaplan, F. (2005). *Les machines apprivoisées*. Paris: Vuibert.
- Kidd, C. D. (2003). *Sociable Robots: The Role of Presence and Task in Human-Robot Interaction*. Unpublished Master Thesis. Massachusetts Institute of Technology, Boston, Massachusetts.
- Koay, K.L., Syrdal, D.S., Ashgari-Oskoei, M., Walters, M.L., & Dautenhahn, K. (2014). Social roles and baseline proxemic preferences for a domestic service robot. *International Journal of Social Robotics*, 6 (4), 469–488.
- Kuchenbrandt, D., & Eyssel, F. (2012). The mental simulation of a humanrobot interaction: Positive effects on attitudes and anxiety toward robots. *Proceedings of the 21st IEEE International Symposium in Robot and Human Interactive Communication (RO-MAN 2012)* (pp. 463–468). Paris, France.
- Lafferty, J.C., & Eady, P.M. (1974). *The desert survival problem*. Plymouth, MI: Experimental Learning Methods.
- Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 46 (1), 50-80.
- Lee, J. J., Knox, W. B., Wormwood, J. B., Beazeal, C., & DeSteno, D. (2013). Computationally modeling interpersonal trust. *Frontiers in Psychology*, 4, 1–14.
- Lee, N., Shin, H., & Sundar, S. S. (2011). Utilitarian vs hedonic robots, role of parasocial tendency and anthropomorphism in shaping user attitudes. *Proceedings of the 6th International Conference on Human-Robot Interaction* (pp. 183–184). Lausanne, Switzerland.
- Li, D., Rau, P. L. P., & Li, Y. (2010). A cross-cultural-study: Effects of robot appearance and task. *International Journal of Social Robotics*, 2 (2), 175–186.
- Lockerd, A., Breazeal, C. (2004). Tutelage and socially guided robot learning. *Proceedings of Intelligent Robots and System*, 4, 3475–3480.
- Lohse, M. (2010). *Investigating the influence of situations and expectations on user behavior – empirical analyses in human-robot interaction*. PhD Dissertation, Bielefeld University.
- Looije, R., Neerinx, M. A., & Cnossen, F. (2010). Persuasive robotic assistant for health self-management of older adults: Design and evaluation of social behaviors. *International Journal of Human-Computer Studies*, 68 (6), 386-397.
- Malle, B. F., Scheutz, M., & Voiklis, J. (2015). Sacrifice One For the Good of Many? People Apply Different Moral Norms to Human and Robot Agents. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction* (pp. 117-124). Portland, USA: ACM.
- Marble, J., L. David, J., Bruemmer, D., Few, A. & Dudenhoefter, D. D. (2004). Evaluation of Supervisory vs. Peer-Peer Interaction with Human-Robot Teams. Paper presented at the *37th Annual Hawaii International Conference on System Sciences*, Hawaii.
- Martelaro, N., Nneji, V., Ju, W., Hinds, P. (2016). Tell Me More: Designing HRI to encourage more trust, disclosure and companionship. Accepted to Human Robot Interaction (HRI 2016). March 7-10, Christchurch, New Zealand.
- Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, 20, 709 –734.
- McBride, M., & Morgan, S. (2010). Trust Calibration for Automated Decision Aids. *Institute for Homeland Security Solutions Research brief*, pp. 1–11. Retrieved from:

- [http://sites.duke.edu/ihss/files/2011/12/McBride\\_Research\\_Brief.pdf](http://sites.duke.edu/ihss/files/2011/12/McBride_Research_Brief.pdf)
- Merritt, S. M., & Ilgen, D. R. (2008). Not all trust is created equal: Dispositional and history-based trust in human-automation interactions. *Human Factors*, 50 (2), 194–210.
- Merritt, S. M., Heimbaugh, H., LaChapell, J., & Lee, D. (2012). I trust it, but I don't know why: Effects of implicit attitudes toward automation on trust in an automated system. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 55 (3), 520–534.
- Monroe, A. E., Dillon, K. D., & Malle, B. F. (2014). Bringing free will down to Earth: People's psychological concept of free will and its role in moral judgment. *Consciousness and cognition*, 27, 100–108.
- Montague, R. P. (2008). Free will. *Current Biology*, 18, R584–R585.
- Mumm, J., & Mutlu, B. (2011). Human-Robot Proxemics: Physical and Psychological Distancing in Human-Robot Interaction. *Proceedings of 6th ACM/IEEE International Conference on Human robot interaction*, (pp. 33–338). Lausanne, Switzerland.
- Nass, C. I., Steuer, J., Tauber, E. R., & Reeder, H. (1993). Anthropomorphism, agency and ethopoeia: computers as social actors. Paper presented at *Computer-Human Interaction Conference (CHI 1993)*. Amsterdam, The Netherlands.
- Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, 56 (1), 81–103.
- Nass, C., Fogg, B. J., & Moon, Y. (1996). Can computers be teammates? *International Journal of Human-Computer Studies*, 45 (6), 669–678.
- Nass, C., Moon, Y., Fogg, B. J., & Reeves, B. (1995). Can computer personalities be human personalities? *International Journal of Human-Computer Studies*, 43 (2), 223–239.
- Natale, L., Nori, F., Metta, G., Fumagalli, M., Ivaldi, S., & Pattacini, U. (2012). The iCub platform: a tool for studying intrinsically motivated learning. In G. Baldassarre, & M. Mirolli (Eds.). *Intrinsically motivated learning in natural and artificial systems*. Berlin: Springer.
- Nichols, S. (2011). Experimental philosophy and the problem of free will. *Science*, 331, 1401–1403.
- Nomura, T. (2014). Influences of Experiences of Robots into Negative Attitudes toward Robots. *Proceedings of the 23rd IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 2014)* (pp. 460–464). Edinburgh, Scotland, UK.
- Nomura, T., Kanda, T., & Suzuki, T. (2006). Experimental investigation into influence of negative attitudes toward robots on human-robot interaction. *AI & Society*, 20, 138–150.
- Oestreicher, L., & Eklundh, K. S. (2006). User Expectations on Human-Robot Co-operation. *Proceedings of the 15th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 2006)* (pp. 91–96). Hatfield, UK.
- Okita, S. Y., Ng-Thow-Hing, V., & Sarvadevabhatla, R. K. (2012). Captain May I? Proxemics Study Examining Factors that Influence Distance between Humanoid Robots, Children, and Adults during Human-Robot Interaction. *Proceedings of the 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, (pp.203-204). Boston, MA.
- Ostrom, T.M. (1969). The relationship between the affective, behavioral, and cognitive components of attitudes. *Journal of Experimental Social Psychology*, 5, 12–30.
- Parasuraman, R. & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, 39, 230–253.
- Park, E., Jenkins, Q., & Jiang, X. (2008). Measuring trust of human operators in new generation rescue robots. In *Proceedings of the JFPS International Symposium on Fluid Power (Vol. 2008, No. 7-2, pp. 489-492)*. The Japan Fluid Power System Society.
- Piaget, J. (1967). *Logique et connaissance scientifique*. Paris: Gallimard.
- Picard, R. W., & Daily, S. B. (2005). Evaluating affective interactions: Alternatives to asking what users feel. In *Proceedings of CHI Workshop 'Evaluating Affective Interfaces'*. Portland, Oregon, USA.
- Popper, K. R. (1991). *La connaissance objective*. Paris: Flammarion.
- Regan, D.T., & Fazio, R. (1977). On the consistency between attitudes and behavior: Look to the method of attitude formation. *Journal of Experimental Social Psychology*, 13 (1), 28–45.
- Riek, L. D. (2012). Wizard of oz studies in hri: a systematic review and new reporting guidelines. *Journal of Human-Robot Interaction*, 1 (1), 119–136.
- Riek, L.D., & Robinson, P. (2008). Robot, rabbit, or red herring? Societal acceptance as a function of classification ease. Paper presented at the *17th International IEEE Symposium on Robot and Human Interactive Communication, Workshop Robots as Social Actors: Evaluating Social Acceptance and Societal Impact of Robotic Agents (RO-MAN 2008)*. Munich, Germany.
- Riether, N. (2013). *On the Profoundness and Preconditions of Social Responses towards Social Robots. Experimental investigations using indirect measurement techniques*. PhD Dissertation, Universität Bielefeld, Germany.
- Rogers, E. M. (1995). *Diffusion of innovations* (4th ed.). New York: Free Press.

- Rotter, J. B. (1971). Generalized expectancies for interpersonal trust. *American Psychologist*, 26, 443–452.
- Salem, M., Lakatos, G., Amirabdollahian, F., & Dautenhahn, K. (2015a). Towards Safe and Trustworthy Social Robots: Ethical Challenges and Practical Issues. In *Social Robotics* (pp. 584-593). Springer International Publishing.
- Salem, M., Lakatos, G., Amirabdollahian, F., Dautenhahn, K. (2015b). Would you trust a (faulty) robot? Effects of error, task type and personality on human-robot cooperation and trust. In: 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI 2015), pp.141-148, Portland, Oregon, USA.
- Salvini, P., Laschi, C., & Dario, P. (2010). Design for Acceptance: Improving Robots' Coexistence in Human Society. *International Journal of Social Robotics*, 2, 451–460.
- Sanders, T., Harpold, B., Kessler, T., & Hancock, P. A. (2015). Interpersonal distance effects on trust relationships in human-robot interaction. In *Proceedings 19th Triennial Congress of the IEA* (Vol. 9, p. 14).
- Schaefer, K.E., (2013). *The Perception And Measurement Of Human-Robot Trust*. PhD Dissertation.
- Schraft, R.D., Meyer, C., Parlitz, C., Helms, E. (2005). Powermate—a safe and intuitive robot assistant for handling and assembly tasks. *IEEE international conference on robotics and automation*, IEEE, Barcelona, pp 4047–4079.
- Serva, M. A., Fuller, M.A., & Mayer, R. C. (2005). “The reciprocal nature of trust: A longitudinal study of interacting teams,” *Journal of organizational behavior*, 26 (6), 625–648.
- Severson, R. L., & Carlson, S. M. (2010). Behaving as or behaving as if? Children's conceptions of personified robots and the emergence of a new ontological category. *Neural Networks, Special Issue on Social Cognition: From Babies to Robots*, 23, 1099–1103.
- Shaw-Garlock, G. (2009). Looking Forward to Sociable Robots. *International Journal of Social Robotics* 1, 249–260.
- Shinozawa, K., Naya, F., Yamato, J., & Kogure, K. (2005). Differences in Effect of Robot and Screen Agent Recommendations on Human Decision-Making. *International Journal of Human-Computer Studies*, 62 (2), 267–279.
- Siegel, M. S. (2008). *Persuasive Robotics How Robots Change our Minds* (Doctoral dissertation, Massachusetts Institute of Technology).
- Sun, H., & Zhang, P. (2006). The role of moderating factors in user technology acceptance. *International Journal of Human-Computer Studies*, 64 (2), 53–78.
- Sung, J-Y., Christensen, H. I., & Grinter, R. E. (2009) “Robots in the Wild: Understanding Long-Term Use” *Proceedings of ACM Conference on Human Robot Interaction (HRI 2009)* (pp. 45-52.). San Diego, California.
- Syrdal, D. S., Koay, K. L., Walters, M. L., & Dautenhahn, K. (2007). A personalized robot companion? - The role of individual differences on spatial preferences in HRI scenarios. *Proceedings of the 16th IEEE International Symposium on Robot and Human interactive Communication (RO-MAN 2007)* (pp. 1143–1148). Jeju, Korea.
- Szalma, J. L., & Taylor, G. S. (2011). Individual differences in response to automation: the five factor model of personality. *Journal of Experimental Psychology: Applied*, 17(2), 71.
- Thrun, S. (2004). Towards a framework for human-robot interaction. *Human Computer Interaction*, 19 (1&2), 9–24.
- Touré-Tillery, M., & McGill, A. L. (2015). Who or what to believe: Trust and the differential persuasiveness of human and anthropomorphized messengers. *Journal of Marketing*, 79(4), 94-110.
- Turkle, S. (2011). *Alone together. Why we expect more from technology and less from each other*. New York: Basic Books.
- van den Brule, R., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., & Haselager, P. (2014). Do Robot Performance and Behavioral Style affect Human Trust? *International Journal of Social Robotics*, 6 (4), 519–531.
- Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27 (3), 425–478.
- Walters, M.L., Dautenhahn, K., Boekhorst, R., Koay, K.L., Kaouri, C., Woods, S., Nehaniv, C., Lee, D., & Werry, I. (2005). The influence of subjects' personality traits on personal spatial zones in a human-robot interaction experiment. *Proceedings of IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN 2005)* (347–352). Nashville, USA.
- Walter-Herrmann, J. & Büching, C. (2014). *FabLab: Of Machines, Makers and Inventors*. Transcript, Verlag.
- Weiss, A., Bernhaupt, R., Lankes (2008). The influence of control on the acceptance of ambient intelligence by elderly people: An explorative study. In Aarts, Emile H.L., Crowley, James L., Ruyter, Boris E. R. de, Gerhäuser, Heinz, Pflaum, Alexander, Schmidt, Janina & Wichert, Reiner (Eds.), *Ambient intelligence – Europeanconference – 2008* (pp. 58–74). Nuremberg, Germany.
- Weiss, A., Bernhaupt, R., Lankes, M., Tscheligi, M. (2009). The USUS Evaluation Framework for Human-Robot Interaction. In *Proceedings of the Symposium on New Frontiers in Human-Robot Interaction* (pp. 158–165). Edinburgh, Scotland, UK.

- Welch, K. C., Lahiri, U., Warren, Z., & Sarkar, N. (2010). An approach to the design of socially acceptable robots for children with autism spectrum disorders. *International Journal of Social Robotics*, 2 (4), 391–403.
- Wullenkord, R., & Eyssel, F. (2014). Improving attitudes towards social robots using imagined contact. Proceedings of the 23rd IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 2014) (pp. 489–49). Edinburgh, Scotland, UK.
- Yagoda, R.E., & Gillan, D.J. (2012). You want me to trust a ROBOT? The development of a human–robot interaction trust scale. *International Journal of Social Robotics*, 4, 235–248.
- Yagoda, R. E. (2011). *What! You want me to trust a robot? The development of a human robot interaction (HRI) trust scale*. Unpublished Master Thesis. North Carolina State University, Raleigh, North Carolina.
- Yanco, H. A., & Drury, J. L. (2002). “A taxonomy for human-robot interaction.” *Proceedings of the AAAI Fall Symposium on Human-Robot Interaction. AAAI Technical Report FS-02-03* (pp. 111–119). Falmouth, MA.
- Young, J. E., Hawkins, R., Sharlin, E., & Igarashi, T. (2009). Toward Acceptable Domestic Robots: Applying Insights from Social Psychology. *International Journal of Social Robotics*, 1, 95–108.
- Yu, A. J. (2015). Decision-Making Tasks. *Encyclopedia of Computational Neuroscience*, 931–937.
- Zaad, L. & Ben Allouch, S. (2008). The Influence of Control on the Acceptance of Ambient Intelligence by Elderly People: An Explorative Study. *Ambient Intelligence, Series Lecture Notes in Computer Science*, 5355, 58–74.