Review Article

The Implicit Sense of Agency: A Decontextualized Indirect Measure of the Sense of Agency Free of Social Desirability

Armand Chatard[1,2]*, Adrien Tedesco[1], Sandrine Belier[3], Olivier Fronty[3] and Ghina Harika-Germaneau[2]

[1]Université de Poitiers, UMR-7295, CNRS, Poitiers, France

[2]Unité de Recherche Clinique, Centre Hospitalier Laborit, Poitiers, France

[3]SBT Human(s) Matter, Paris, France

Corresponding author: Armand Chatard, Centre de Recherches sur la Cognition et l'Apprentissage, CNRS 7295, MSHS, Bat A5, 5 rue Théodore Lefebvre, TSA 21103, 86073 Poitiers CEDEX, France

Abstract

The sense of agency (SoA) refers to the feeling of initiating and controlling one's actions. SoA is thought to play an important role in a number of mental illnesses. SoA can be assessed with both explicit (self-report) and implicit (perceptual judgment) measures. However, existing measures tap only specific aspects of the SoA (e.g., temporal binding). To date, there is lack of general (decontextualized) measure of implicit SoA. The aim of the present research was to develop a new decontextualized measure of implicit SoA, based on a reaction time interference paradigm (a variant of the semantic Simon task). An examination of its relation with a decontextualized explicit measure of the SoA provided evidence for its convergent and discriminant validity. In Study 1 (N = 115 French citizens), there was a significant modest correlation between the direct and indirect measures of the SoA. In Study 2 (N = 101 British citizens), this finding was replicated in an English-speaking sample. Moreover, it was found that the indirect, but not the direct, measure of the SoA was free of social desirability bias. These findings suggest that the decontextualized direct and indirect measures of the SoA used in the present work assess the same latent underlying construct. The indirect measure may be a useful tool to assess the SoA whenever social desirability is a concern.

Keywords: Sense of agency, Direct and indirect measures, Construct validity, Reaction time-based interference measure, Social desirability

The Implicit Sense of Agency: A Decontextualized Indirect Measure of the Sense of Agency Free of Social Desirability

The sense of agency (SoA) refers to the feeling of initiating and controlling one's own actions [1,2], and is believed to play a central role in a variety of normal and pathological behaviors [3]. Over the last two decades, researchers have been increasingly interested in exploring the SoA, and both direct and indirect measures of the construct have been developed [4]. Direct measures include self-rating scales and self-report questionnaires, where individuals are asked to make judgments about their experience of agency. Indirect measures typically include performance on perceptual tasks that require discriminating and evaluating self-generated and externally generated stimuli. These measures assess specific aspects of the SoA, and are often uncorrelated to each other [5-7]. Recently, a more general self-report measure of the SoA has been developed [6]. This direct "decontextualized" measure assesses general feelings about the SoA in general, rather than on a specific task. To date, however, there is a lack of "decontextualized" indirect measure of the SoA. The present study aimed to redress this neglect by developing a new indirect measure of implicit SoA, semantically related to the explicit judgment of agency, but with an important advantage over direct

measures: control over social desirability bias [8]. In what follows, extant direct and indirect measures are described, before addressing the problem of social desirability.

Indirect Measures of the Sense of Agency

The SoA has classically been assessed in the laboratory with specific tasks such as the temporal binding task [9] and the sensory attenuation task [10]. These tasks provide an indirect measure of the sense of agency, as the participants are not explicitly asked to make judgment about their SoA. Rather, their SoA is inferred from their evaluations of specific attributes of self-generated and externally generated stimuli. In the temporal binding task, the participant has to estimate the time delay between two stimuli (two tones) when the first stimulus is either triggered by one's own action (a key press) or externally generated (initiated by a computer program). Temporal binding refers to a reduction of the perceived time delay when the participant is the initiator of the action, compared to when the action is externally generated. In the sensory attenuation task, the participant has to evaluate the intensity of a self-initiated action, compared to the intensity of a computer-generated action. The sensory attenuation effect refers to a reduction in the subjective intensity of a self-initiated action, compared to the intensity of the same action initiated by another agent. This sensory attenuation phenomenon has been found

for a variety of stimuli, including haptic, auditory and visual stimuli [11-15]. Even if there is still controversy about the precise underlying mechanisms, temporal binding and sensory attenuation have been used as SoA proxies in numerous studies [16-20].

Direct Measures of the Sense of Agency

According to a prominent model of the SoA [21] the SoA does not consist only of sensorimotor processes but also involves higher-order conceptual processes. At the most basic level, a perceptual representation (the feeling of agency) is formed based on an integration of proprioception, sensory feedback, and feed-forwards cues. However, the feeling of agency is further elaborated at a higher-order level to form a conceptual representation (the judgment of agency), contingent on current thoughts, goals, social influence and contextual cues. Therefore, the two-step account of agency [21], suggests that the SoA reflects a complex integration of two interrelated representations, the feeling of agency, at the perceptual level, and the judgment of agency, at the conceptual level.

The judgment of agency has often been assessed in specific perceptual tasks by asking participants to attribute a cause to an action-effect [22]. For example, in the temporary binding and sensory attenuation tasks participants are often asked to rate the degree to which they think that they are personally responsible for a particular action-effect on a scale from "not at all" to "absolutely" [23-26]. More recently, researchers have also sought to develop less contextualized measures of the judgment of agency, unrelated to a particular task. The sense of agency scale is a self-report questionnaire assessing the degree to which people think that they are personally responsible for what happens in their life. The scale includes items such as "I am completely responsible for everything that results from my actions" and consists of two main factors: judgment of positive (e.g. "Things I do are subject only to my free will") and negative (e.g. "My actions just happen without my intention") agency. To date, this self-report scale has been validated in two languages: Hebrew and French [27]. Tapal et al. reported a moderate-to-strong correlation (r = 0.35) between the negative subscale of the judgment of agency and obsessive-compulsive symptoms, providing preliminary evidence for the predictive validity of this direct measure.

Correlations between Direct and Indirect Measures of the Sense of Agency

Although people are generally aware of being the authors of their own actions, sometimes they underestimate or overestimate their agency (even in the absence of pathological disorder). The idea that the sense of agency is sometimes "just an illusion" is not new [28,29]. It is supported by a number of experiments using subliminal priming procedures [30-32]. Also consistent with the notion that individuals may lack insight on the cause of their actions, previous studies have found no significant relations between direct and indirect measures of the SoA. For example, Dewey and Knoblich measured the SoA using two indirect measures (the temporal binding task and the sensory attenuation task) in a sample of 78 young adults. They also assessed explicit judgment of agency in these specific tasks using self-rating scales. Although the sensory attenuation and temporal binding

effects were replicated in these studies, somewhat surprisingly, there were no significant correlations between direct and indirect measures of the SoA. Neither temporal binding nor sensory attenuation was significantly correlated with explicit judgment of agency. These findings suggest that, even in the context of a simple perceptual task, individuals are unable to explicitly discriminate the actions they are responsible for to those they are not. In other words, these findings suggest that individuals are "strangers to their own actions".

The findings reported by Dewey and Knoblich may indicate that the feeling and judgment of agency are two completely distinct constructs that do not overlap. However, this is at odds with the two-step account of agency which predicts that the judgment of agency is mainly a cognitive elaboration of the feeling of agency. More generally, recent research and theorizing indicate that direct and indirect measures of cognition often assess a single underlying process rather than two fundamentally distinct processes [33]. Therefore, there are theoretical reasons to expect the feeling and judgment of agency to be positively correlated with each other.

The Role of Social Desirability in the Link between Direct and Indirect SoA Measures

One variable that may affect the relation between direct and indirect measures of the SoA is social desirability. Individuals may be motivated to exaggerate their judgment of agency to give a good impression to others (impression management) or to self-enhance (self-deceptive enhancement) [34,35]. Direct and indirect measures are differently affected by social desirability responding. Direct measures are often biased by self-presentation or social desirability, because it is fairly obvious what is being measured by self-reports, and because responses on such measures are easily controllable. In contrast, indirect measures are more resistant, or less vulnerable, to social desirability, since what is being measured by these tasks is less obvious, and responses are less controllable. As a consequence, correlations between direct and indirect measures of the same construct are often moderated by social desirability, such that correlations decrease as a function of increasing social desirability. For example, social desirability moderates the relation between direct and indirect measures of attitudes towards asylum seekers. Using multilevel modelling analyses, Nosek [36-39] reported similar findings for 57 different attitude objects. Therefore, social desirability is one factor that is likely to affect the correlation between direct and indirect measures of the SoA. To the best of our knowledge, no study on the SoA has sought to control for social desirability responding.

The Present Studies

The aim of the present studies was twofold. First, a new indirect measure was developed to assess decontextualized SoA. Second, the relation between direct and indirect measures of the SoA was tested with and without taking into account the degree to which individuals are concerned with social desirability, to better identify the boundary conditions for these correlations to emerge. Correlations between direct and indirect measures of the SoA would provide support for a core prediction of the two-step account of agency. If the SoA results from the interplay of both motor and conceptual representations, then

a positive correlation between direct and indirect decontextualized measures of the SoA is to be expected. Understanding under which precise conditions these correlations emerge is also important to inform researchers about when and why it is preferable to use indirect versus direct SoA measures.

Previous studies in which correlations between direct and indirect measures of the SoA were tested have focused on specific perceptual events – they were highly contextualized. In an effort to extend this work to a more general and decontextualized context, we used the newly developed SoA scale as a direct measure. To date however, there is no equivalent, decontextualized indirect measure in the literature. Therefore, in Study 1, we developed an indirect decontextualized SoA measure. To do this, a reaction time-based interference paradigm was used. This indirect measure was modeled after a measure previously used to assess implicit feeling of self-control [40], see also [41]. Study 1 tested the correlations between the direct and indirect measures of the SoA.

Study 2 aimed to replicate Study 1 findings. Another aim was to test correlations between direct and indirect measures of the SoA and social desirability. It was predicted that social desirability would be significantly correlated with the direct measure of the SoA. Such a correlation would be consistent with the idea that individuals do not accurately (e.g., exaggerate) report their SoA on direct measures. Finally, we examined whether correlations between direct and indirect measures of the SoA are moderated by social desirability, such that they decrease as a function of increasing social desirability. Such moderation would provide new insight on when and why direct and indirect SoA measures are correlated.

Study 1

In Study 1, we tested for the first time whether individual differences in direct and indirect measures of the SoA are correlated when using decontextualized measures, rather than direct and indirect measures specific to predetermined perceptual events.

Method

Participants

Because the study was conducted during the Covid-19 pandemic and the lockdown in France, it was run on the Internet. One hundred and fifteen participants (41 females, 73 males, and 1 "other", mean age = 28.84, SD = 9.11) were recruited online from a crowdsourcing platform designed for scientific research (Prolific Academic, https://www.prolific.co/). Only French nationals fluent in French were eligible for the study (inclusion criteria). Participants received payment of 2.50£. All participants read and completed a consent form for online research. The study was approved by the Ethical Review Committee for Research of the University of Tours and Poitiers (CER-TP).

Indirect Measure of the SoA

As an indirect measure of the SoA, we used a variant of the semantic Simon task [42]. In previous studies, this task has been successfully used to measure implicit feelings of self-control. In the present study, this task was adapted to measure implicit feelings of agency. The task is based on a reaction time interference paradigm. In each trial, participants were asked to indicate as fast as possible whether the stimulus (series of words) displayed on the screen was written in uppercase or in lowercase letters by pressing the appropriate key on the keyboard (I = lowercase and E = uppercase). The stimuli (e.g., "INITIATING AN ACTION") were directly taken from the items of the SoA scale. Participants were explicitly asked to ignore the meaning of the stimuli and to respond only to its perceptual form (lowercase vs. uppercase letters). The stimuli described an action or a state of mind. Each trial was preceded by a prime: "Me" or "Others", presented for 1500 ms. After a fixation cross of 500 ms, the stimulus was displayed. Participants were asked to disregard the primes, described as distractors.

The task comprised twelve stimuli: six series of words congruent with the SoA (e.g., free will, responsible for my action) and six series of words incongruent with the SoA (e.g., surprised by my actions, behaving like a robot). Each stimulus was presented four times with the prime "Me" and four times with the prime "Others", equally often in uppercase and lowercase letters, resulting in a total of 96 trials. The stimuli were presented in a full random order.

Direct Measure of the SoA

As a direct measure of the SoA, participants completed the SoA scale. This self-report measure consists of 13 items (e.g., "Things I do are subject only to my free will", α = 0.79). Participants reported their degree of agreement with each item on a 7-point Likert scale (1 = totally disagree, and 7 = totally agree). The mean of the 13 items was computed, with higher scores indicating greater judgment of agency (M = 5.08, SD = 0.78). The SoA scale has two subscales: the sense of positive agency (SoPA, e.g., "Things I do are subject only to my free will", α = 0.70) and the sense of negative agency (SoNA, e.g., "Nothing I do is actually voluntary", α = 0.71). High scores on the SoPA indicate explicit judgment of positive agency (M = 4.98, SD = 0.93), whereas high scores on the SoNA indicate explicit judgment of negative agency (M = 2.82, SD = 0.87).

Procedure

The script of the experiment was written in PsychoPy3 (https://www.psychopy.org/), and them converted to JavaScript using Pavlovia (https://pavlovia.org/). The link to the study was then posted on the Prolific Academic platform for participant recruitment. Participants first performed eight practice trials of the indirect measure of the SoA, with agency-unrelated stimuli. Then, they completed the indirect measure of the SoA. This task took about 5 minutes to complete. Finally, participants completed the direct measure of the SoA. At the end of the study, they were thanked and debriefed.

Results

Data Cleaning

Concerning the indirect measure data, reaction times (RTs) larger than 1500ms (deemed to be too long) as well as those associated with incorrect responses were excluded (an average of 15.27% of the trials, SD = 11.80). Average RTs on schema-congruent trials preceded by the "Me" prime (M = 603.21, SD = 121.85) and preceded by the "Others"

prime (M = 602.83, SD = 123.89) were then computed. In the same way, the average RTs on schema-incongruent trials preceded by the "Me" prime (M = 600.86, SD = 130.62) and preceded by the "Others" prime (M = 602.53, SD = 126.18) were computed. As in previous studies [40,41], we calculated a priming effect for schema-congruent trials (RTs on Others-congruent trials minus RTs on Me-congruent trials) and for schema-incongruent trials (RTs on Me-incongruent trials minus RTs on Others-incongruent trials). Higher scores on both of these indicators reflect higher agency. Thus, to create an overall SoA score, the priming effect on schema-congruent and schema-incongruent trials (M = 4.19, SD = 64.67) were added. Higher scores on this measure thus indicate higher semantic interference of schema-congruent rather than schema-incongruent stimuli with motor responses based on perceptual judgments. In other words, positive scores on this indirect measure indicate an association between the self (vs. others) with an implicit feeling of positive (vs. negative) agency.

Main Analysis

After data cleaning, the distribution of SoA scores on the indirect measure followed a normal distribution, W = 0.98, p = 0.088, Skew = 0.39, SE = 0.22. In the same way, the SoA scores on the direct measure were normally distributed, W = 0.97, p = 0.054, with a slightly left-skewed tail, Skew = −.44, SE = 0.22. Thus, parametric Pearson correlations are reported to examine the associations between direct and indirect measures of the SoA.

The correlation coefficients are reported in Table 1. As shown in this table, there was a significant positive correlation between the direct and indirect SoA measures (r = 0.24, p < 0.01). This correlation was the most pronounced with the negative subscale of the direct measure. The correlation between the implicit SoA and the SoNA was moderate-to-strong (r = −.32, p < 0.001). In contrast, the correlation between the implicit SoA and the positive subscale of the direct measure was not significant (r = 0.09) (Table 1).

Supplementary Analysis

To gain further insight on how the negative subscale of the direct SoA measure related to the interference effect in the indirect measure further analyses were conducted. The mean RTs on the indirect measure were submitted to a mixed ANOVA with schema congruence (congruence vs. incongruence) and prime (me vs. others) as within-subjects factors, and the SoNA scores (low or high based on a median split) as a between-subjects factor. In this analysis, the three-way interaction was significant, F(1, 111) = 15.00, p < 0.001, η^2 = 0.11. The means indicated that participants with low scores on the SoNA subscale (i.e., participants high in the SoA) showed an interference effect (Figure 1 for the means). They showed shorter RTs on schema-congruent trials when primed with "Me" rather than "Others", and larger RTs on schema-incongruent trials when primed with "Me" rather than "Others". In contrast, participants with high scores on the SoNA subscale (i.e., participants low in the SoA) showed a reverse-interference effect.

Discussion

In this study, the SoA was measured with a direct and a newly developed indirect measure. The two measures assessed decontextualized SoA, rather than task- or domain-specific SoA. The direct measure was a validated self-report questionnaire. The indirect measure was based on a reaction time-based interference paradigm. There was a significant correlation between the direct and indirect measures of the SoA, consistent with the view that the two measures tap into overlapping processes. An unexpected finding was that the correlation was especially pronounced for the negative subscale of the direct measure. Study 2 aimed to replicate and extend these findings and to examine the relation between direct and indirect measures of the SoA, on one hand, and social desirability, on the other hand.

Study 2

The aim of Study 2 was twofold. The first aim was to replicate findings from Study 1 in an English-speaking sample, rather than a French-speaking sample. The second aim was to extend Study 1 findings by testing the hypothesis that the direct, but not the indirect measure of the SoA is positively related to social desirability. It was predicted that social desirability would moderate the association between direct and indirect measures of the SoA, such that relation between direct and indirect measures would decrease as a function of increasing social desirability. In other words, the direct measure of the SoA should be related to the indirect measure of the SoA only or mainly when social desirability is relatively low.

Table 1: Correlation coefficients between the direct and indirect measures of the SoA in Study 1.

	Implicit SoA	Explicit SoA		SoPA		SoNA	
Implicit SoA	—	0.246	**	0.097		−0.320	***
	—	0.009		0.307		<.001	
Explicit SoA		—		0.857	***	−0.882	***
		—		<.001		<.001	
SoPA				—		−0.513	***
				—		<.001	
SoNA						—	
						—	

Note. Entries are Pearson's correlation coefficients (top line) and exact p-values (bottom line).
Implicit SoA: Sense of agency on the indirect measure, Explicit SoA: Sense of agency on the direct measure, SoPA: Self-reported Sense of Positive Agency, SoNA: Self-reported Sense of Negative Agency *p < 0.05, **p < 0.01, ***p < 0.001
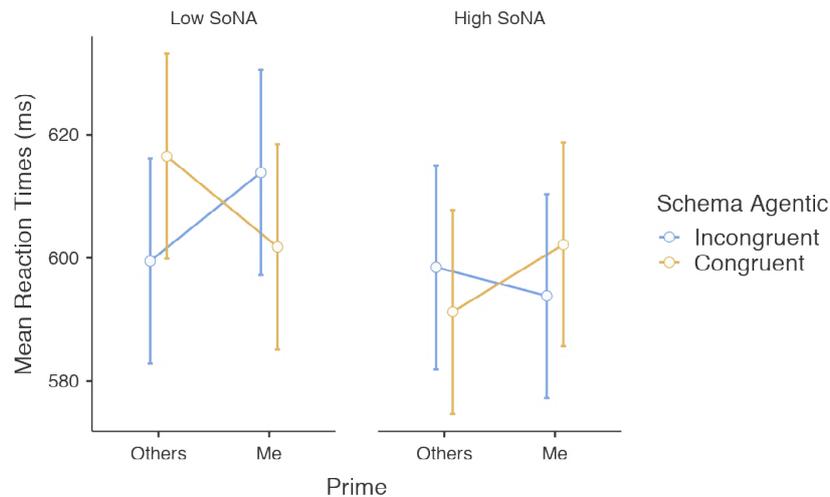
Figure 1: Means of reaction times in the indirect measure of the SoA as a function of the direct measure (SoNA).

Method

Participants

A sample of 101 British participants was recruited on the Prolific Academic platform (58 females and 43 males, mean age = 31.48 years, SD = 13.29). The script of the experiment was written in English. There were no sound theoretical reasons to expect differences between the French and the UK sample. This study might thus contribute to extend the generality of Study 1 findings.

The sample size was large enough to provide adequate power $(1 – β = 0.80)$ to detect a correlation in the same direction and as large as the one found in Study 1 $(r = 0.246$, with $α = 0.05$). Only UK nationals fluent in English were eligible for the study. Participants received a payment of 2.50£. All participants read and completed a consent form for online research. The study was approved by the Ethical Review Committee for Research of the University of Tours and Poitiers (CER-TP).

Materials

Indirect Measure of the SoA

The indirect measure of the SoA developed in Study 1 was translated into English. The data were cleaned, and the scores on the indirect measure of the SoA were computed as in Study 1 (M = 7.03, SD = 70.91).

Direct Measure of the SoA

The English version of the SoA scale reported in Tapal et al. [6] was used. As in Study 1, the 13 items of the scale were averaged to form a composite score of the explicit SoA $(α = 0.70, M = 4.91, SD = 0.768)$. The means of the two subscales, the SoPA $(α = 0.70, M = 4.81, SD = 0.99)$ and the SoNA $(α = 0.70, M = 3.00, SD = 0.95)$, were also computed.

Social desirability

Social desirability responding was assessed with the Balanced Inventory of Desirable Responding short form (BIDR-16, [43]).

This 16-item questionnaire allows to assess two main components of social desirability: self-deceptive enhancement (the tendency to give self-reports that are honest but positively biased) and impression management (deliberate self-presentation to an audience). The 16 items were averaged to form a global score of social desirability responding $(α = 0.76, M = 3.96, SD = 0.77)$. Average scores were also calculated for self-deceptive enhancement $(8$ items, $α = 0.73, M = 3.80, SD = 0.94)$ and impression management $(8$ items, $α = 0.69, M = 4.12, SD = 0.96)$.

Procedure

The procedure was highly similar to the one used in Study 1, except that the study was in English and that the participants completed the BIDR-16 scale after the indirect and direct SoA measures.

Results

The SoA scores on the direct measure were normally distributed, $W = 0.99, p = 0.70$, Skew = $– 0.11$, SE = 0.24. In this sample, however, the SoA scores on the indirect measure did not follow a normal distribution, $W = 0.88, p < 0.001$, Skew = 1.30, SE = 0.24. A log transformation was thus applied to the data. However, the p-values and Pearson correlation coefficients were very similar when using the transformed and the non-transformed scores. The findings were also very similar when using Spearman rather than Pearson correlation coefficients. Thus, to be consistent with Study 1 and to avoid redundancy, only the results of Pearson correlation coefficients obtained with the non-transformed scores are reported in what follows.

Replication of Study 1 Findings

Table 2 presents correlations between the direct and indirect SoA measures. As shown in this table, there was a significant positive correlation between the direct and indirect measure of the SoA $(r = 0.29, p < 0.005)$. Also, as in Study 1, the implicit SoA was significantly correlated with the SoNA $(r = -.36, p < 0.001)$, but not with the SoPA $(r = 0.08)$. These results thus replicated findings from Study 1 (Table 2).

Table 2: Correlations between the direct and indirect measures of the SoA and social desirability scores (Study 2).

	Implicit SoA		Explicit SoA		SoNA		SoPA		Desirability		SDE		IM	
Implicit SoA	—		0.292	**	−0.363	***	0.080		−0.089		−0.070		−0.076	
	—		0.004		<.001		0.438		0.387		0.499		0.461	
Explicit SoA			—		−0.808	***	0.752	***	0.228	*	0.241	*	0.130	
			—		<.001		<.001		0.022		0.015		0.196	
SoNA					—		−0.220	*	−0.124		−0.106		−0.094	
					—		0.027		0.218		0.290		0.348	
SoPA							—		0.239	*	0.279	**	0.109	
							—		0.016		0.005		0.276	
Desirability									—		0.808	***	0.815	***
									—		<.001		<.001	
SDE											—		0.318	**
											—		0.001	
IM													—	
													—	

Note. Entries are Pearson's correlation coefficients (top line) and exact p-values (bottom line). Implicit SoA: Sense of agency (indirect measure), Explicit SoA: Sense of agency (direct measure), SoPA: Sense of Positive Agency (direct measure), SoNA: Sense of Negative Agency (direct measure), Desirability: Social desirability, SDE: Self-Deceptive Enhancement, IM: Impression Management, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$
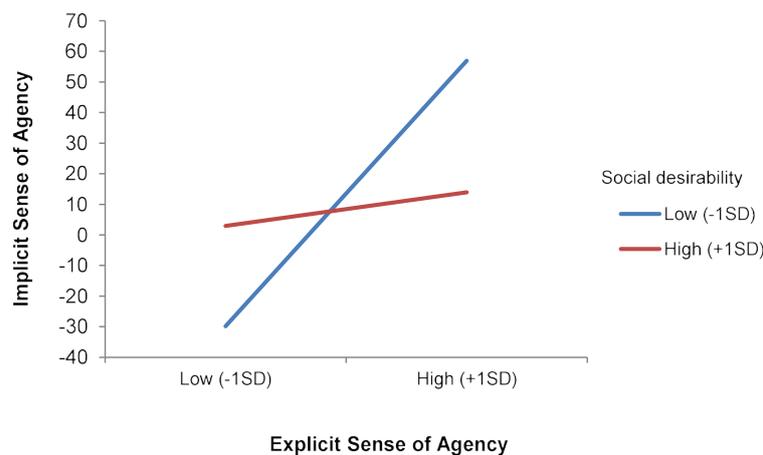

Figure 2: Predicted means of implicit sense of agency.

As in Study 1, further analyses were conducted to gain insight on how the negative subscale of the direct measure of the SoA was related to the interference effect in the indirect measure. The mean RTs on the indirect measure were submitted to a mixed ANOVA with schema congruence and prime as within-subjects factors and the SoNA scores (low or high based on a median split) as a between-subjects factor. In this analysis, the three-way interaction was significant, $F(1, 94) = 6.26$, $p = 0.014$, $\eta^2 = 0.062$. The means followed the same pattern as the one found in Study 1.

Relations between SoA and Social Desirability

Another important aim of this study was to examine the correlations between the direct and indirect measures of the SoA and social desirability scores (Table 2). As expected, the indirect measure of the SoA was not related to social desirability scores (all ps > 0.30). In contrast, the direct measure of the SoA was significantly positively correlated with social desirability in general ($r = 0.22$, $p = 0.022$), and with self-deceptive enhancement ($r = 0.24$, $p = 0.015$), in particular. Further analyses showed, however, that the two subscales of the explicit SoA were differently related to social desirability scores. While the SoNA was not associated with social desirability scores (all ps > 0.20), the SoPA was significantly associated with social desirability ($r = 0.24$, $p = 0.016$) in general, and with self-deceptive enhancement ($r = 0.27$, $p = 0.005$), in particular.

Finally, the moderation of the relation between direct and indirect measures of the SoA by social desirability was assessed. Implicit SoA was regressed on social desirability, explicit SoA, and the product term between social desirability and explicit SoA. Consistent with a moderation hypothesis, the product term was significant, $B = −32.09$, $SE = 11.15$, $t = −2.87$, $p < 0.005$. This interaction showed that the relation between explicit and implicit SoA decreased as a function of increasing social desirability (Figure 2 for predicted means). Simple slope analyses confirmed that the relation between explicit and implicit SoA was significant at low level (−1SD) of social desirability, $B = 56.99$, $SE = 12.89$, $t = 4.41$, $p < 0.001$, but not at high level (+1SD) of social desirability, $B = 7.24$, $SE = 11.84$, $t = 0.61$, $p = 0.54$.

Discussion

In this study, the findings observed in Study 1 were replicated. The direct and indirect measures of the SoA were significantly correlated, and this correlation was larger for the negative than for the positive subscale of the direct measure. Evidence was also found that the direct but not the indirect SoA measure is contaminated by social desirability bias. More particularly, social desirability seems to be positively related to self-reports on the positive subscale of the SoA. Finally, results of moderation analyses confirmed that social desirability suppressed the relation between the direct and indirect SoA measures.

General Discussion

Research and theory in cognitive psychology suggest that the SoA arises from a complex interplay between sensorimotor and cognitive (conceptual) processes [44]. Somewhat surprisingly, however, researchers have consistently failed to find correlations between indirect measures (perceptual tasks) and direct measures (explicit evaluation) of the SoA. This is inconsistent with the view that the explicit judgment of agency is intrinsically linked to the implicit feeling of agency. That being said, research conducted so far has focused on specific experimental tasks, and little was known about the correlation between direct and indirect measures of the SoA in less specific and constrained situations. In the present studies, this issue was directly addressed.

Study 1 relied on a semantic Simon task to measure the SoA at the implicit level, and a self-report scale to assess the SoA at the explicit level. Performance on the indirect task depended on both sensorimotor and semantic processes. While participants were instructed to ignore conceptual information (the meaning of the stimuli) and to focus on perceptual information (the shape of the stimuli), they were unable to do so. More precisely, they showed a facilitation effect when the conceptual information was consistent with their explicit judgment of agency, and an inhibition effect when the conceptual information was inconsistent with their explicit judgment of agency. The higher SoA they reported at the explicit level, the stronger interference (or facilitation) effect they showed in the indirect task. These findings strongly suggest that there is an overlap between conceptual and sensorimotor processes in the experience of agency.

In Study 2, these findings were replicated and significant correlation between social desirability and the direct SoA measure was found. The higher levels of social desirability in general and of self-deceptive enhancement in particular the participants reported, and the higher were their explicit judgments of agency. In addition, social desirability moderated the strength of the association between direct and indirect measures of the SoA. When social desirability was relatively low rather than high, the implicit feeling of agency was strongly linked with the explicit judgment of agency. These findings might contribute to explain why previous studies have failed to find correlations between direct and indirect measures of the SoA. Social desirability can also contribute to explain why in the present studies the SoPA failed to predict the implicit feeling of agency (the indirect measure). As observed in the present research, the two subscales of the explicit measure are differently related to social desirability responding.

The SoPA is contaminated by social desirability responding, whereas the SoNA is not (or is less) biased.

Theoretical and Practical Implications

The present findings have a number of theoretical, methodological, and practical implications. From a theoretical perspective, the reported results are in line with the two-step account of the SoA. According to this model, the explicit judgment of agency is a direct consequence of the implicit feeling of agency. Therefore, the two-step account of the SoA would logically predict correlations between direct and indirect measures of the SoA. In line with this model, significant correlations between an indirect measure of the SoA (a reaction time-based interference task), and a direct measure of the SoA (a self-report scale) were found in two studies. Boundary conditions for these correlations were also identified in relation to social desirability concerns. The explicit judgment of agency is more likely to reflect the implicit feeling of agency when social desirability is low, rather than high. Taken together, these findings complement and extend previous findings, and offer a finer-grained analysis of the relation between different aspects of the SoA.

It might be argued that the sematic Simon task overestimates the correlations between the feeling and judgment of agency because it is not a "process pure" task. Indeed, this reaction time interference paradigm, like most other cognitive tasks (e.g., the Stroop task), is not "process pure" [45]. It involves a complex interaction between semantic/conceptual and perceptual/motor processes. However, this does not mean that it fails to capture the implicit feeling of agency. Indeed, behavioral and neuroimaging studies suggest that multiple self-knowledge systems exist (evidence-based and intuition-based, see [46]). The intuitive sense of the self does not require conscious (or deliberate) reflection. Thus, it might be argued that the semantic Simon task is well suited to study the implicit feeling of agency because performance on this task reflects the interference of the automatic activation of intuition-based self-knowledge with perceptual and motor processing.

From a methodological and practical perspective, the present findings have important implications for the assessment of the SoA. Clearly, the simplest way to measure the SoA is to ask direct questions about various aspects of a person's sense of agency. However, it may not be the best way to measure the SoA because self-reports are sensitive to social desirability responding and self-presentation strategies. Here, it was shown that individuals tend to overestimate their SoA on a direct, self-report scale because of social desirability. The present studies were conducted online, in private and anonymous conditions, which are likely to minimize the problem of social desirability. Social desirability may be an even bigger problem in more ecological conditions (when responses are not anonymous or when they are public). Importantly, however, it was found that the indirect measure of the SoA was free of social desirability bias. The indirect measure showed very good convergent validity with the direct measure, but only when social sociability was low. This suggests that the indirect measure of the SoA should be favored whenever social desirability is a concern.

Limitations

There are a number of limitations of the present studies that deserve to be addressed in future research. First, the indirect measure used in the present work needs to be further validated. The present findings provide evidence for the convergent and discriminant validity of the indirect measure with the direct measure in two languages. However, the discriminant validity of the indirect measure, compared to the direct measure, needs to be further explored. For example, Tapal et al. found a moderate-to-strong correlation between the SoNA subscale and obsessive-compulsive symptoms. In future studies, it would be interesting to examine the respective associations of the direct and indirect measures of the SoA with obsessive-compulsive symptoms. It might contribute to better understand how the SoA is related to clinical disorders.

Another limitation is that there were no behavioral measures in the present studies to test the predictive validity of the indirect measure of the SoA. Further studies are needed to examine the predictive validity of the indirect measure for a variety of criterion outcomes. Interestingly, Huntjens et al. used a similar indirect measure to assess trait self-control and showed that this indirect measure has superior predictive power for spontaneous trait-related behavior (task persistence and delay discounting) than direct self-report measures of self-control. These findings should be taken with caution, as the sample size was quite small in their study. However, Huntjens et al.'s findings clearly suggest that the indirect measure of the SoA has the potential to predict relevant behaviors, and it should be tested in subsequent lab studies.

Conclusion

The clear take-home message of the present work is that implicit SoA can be measured with a decontextualized reaction time interference paradigm in less than 5 minutes. The task is currently available in two languages (French and English) and is free of social desirability bias. The implicit feeling of agency, as measured with this task, is positively related (with moderate-large effect size) to more explicit judgment of agency, but only when social desirability is low. Thus, whenever social desirability is a concern, the indirect measure might provide a better estimate of the SoA than a self-report scale.

References

1. Haggard P, Chambon V (2012) Sense of agency. *Current Biology* 22: 390-392.

2. Synofzik M, Vosgerau G, Voss M (2013) The experience of agency: an interplay between prediction and postdiction. *Frontiers in Psychology* 4: 127. [crossref]

3. Moore JW (2016) What is the sense of agency and why does it matter? *Frontiers in Psychology* 7: 1272. [crossref]

4. Moore JW, Middleton D, Haggard P, Fletcher PC (2012) Exploring implicit and explicit aspects of sense of agency. *Consciousness and Cognition* 21: 1748-53. [crossref]

5. Dewey JA, Knoblich G (2014) Do implicit and explicit measures of the sense of agency measure the same thing? *PloS One* 9: e110118.

6. Tapal A, Oren E, Dar R, Eitam B.(2017) The sense of agency scale: A measure of consciously perceived control over one's mind, body, and the immediate environment. *Frontiers in Psychology* 8: 1552. [crossref]

7. Merrick S.(1995) Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *Am Psychologist* 50: 741.

8. Gawronski B, LeBel EP, Peters KR (2007) What do implicit measures tell us?: Scrutinizing the validity of three common assumptions. *Perspectives on Psychological Science* 2: 181-93. [crossref]

9. Haggard P, Clark S, Kalogeras J (2002) Voluntary action and conscious awareness. *Nature Neuroscience* 5: 382-5.

10. Desantis A, Weiss C, Schütz-Bosbach S, Waszak F (2012) Believing and perceiving: authorship belief modulates sensory attenuation. *PLoS One* 7: e37959. [crossref]

11. Blakemore SJ, Wolpert DM, Frith CD (1998) Central cancellation of self-produced tickle sensation. *Nature Neuroscience* 1: 635-40. [crossref]

12. Blakemore SJ, Frith CD, Wolpert DM (1999) Spatio-temporal prediction modulates the perception of self-produced stimuli. *Journal of Cognitive Neuroscience* 11: 551-9. [crossref]

13. Weiss C, Herwig A, Schütz-Bosbach S (2011) The self in action effects: selective attenuation of self-generated sounds. *Cognition* 121: 207-18. [crossref]

14. Reznik D, Henkin Y, Levy O, Mukamel R (2015) Perceived loudness of self-generated sounds is differentially modified by expected sound intensity. *PloS One* 10: e0127651. [crossref]

15. Gentsch A, Schütz-Bosbach S (2011) I did it: unconscious expectation of sensory consequences modulates the experience of self-agency and its functional signature. *Journal of Cognitive Neuroscience* 23: 3817-28. [crossref]

16. Engbert K, Wohlschläger A, Haggard P (2008) Who is causing what? The sense of agency is relational and efferent-triggered. *Cognition* 107: 693-704. [crossref]

17. Desantis A, Roussel C, Waszak F (2011) On the influence of causal beliefs on the feeling of agency. *Consciousness and Cognition* 20: 1211-20.

18. Haering C, Kiesel A (2012) Mine is earlier than yours: Causal beliefs influence the perceived time of action effects. *Frontiers in Psychology* 3: 393. [crossref]

19. Moore JW, Dickinson A, Fletcher PC (2011) Sense of agency, associative learning, and schizotypy. *Consciousness and Cognition* 20: 792-800. [crossref]

20. Moore JW, Wegner DM, Haggard P (2009) Modulating the sense of agency with external cues. *Consciousness and Cognition* 18: 1056-64. [crossref]

21. Synofzik M, Vosgerau G, Newen A (2008) Beyond the comparator model: a multifactorial two-step account of agency. *Consciousness and Cognition* 17: 219-39.

22. Wegner DM, Wheatley T (1999) Apparent mental causation: Sources of the experience of will. *American Psychologist* 54: 480. [crossref]

23. Aarts H, Custers R, Wegner DM (2005) On the inference of personal authorship: Enhancing experienced agency by priming effect information. *Consciousness and Cognition* 14: 439-58.

24. Dewey JA, Carr TH (2013) When dyads act in parallel, a sense of agency for the auditory consequences depends on the order of the actions. *Consciousness and Cognition* 22: 155-66. [crossref]

25. Frith C (2013) The psychology of volition. *Experimental Brain Research* 229: 289-99. [crossref]

26. Sato A, Yasuda A (2005) Illusion of sense of self-agency: discrepancy between the predicted and actual sensory consequences of actions modulates the sense of self-agency, but not the sense of self-ownership. *Cognition* 94: 241-55.

27. Hurault JC, Broc G, Crône L, Tedesco A, Brunel L (2020) Measuring the Sense of Agency: A French Adaptation and Validation of the Sense of Agency Scale (F-SoAS). *Frontiers in Psychology* 11. [crossref]

28. Wegner DM (2002) The illusion of conscious will. *MIT press*.

29. Wegner DM (2003) The mind's best trick: How we experience conscious will. *Trends in Cognitive Sciences* 7: 65-9. [crossref]

30. Aarts H, Custers R, Marien H (2009) Priming and authorship ascription: When nonconscious goals turn into conscious experiences of self-agency. *Journal of Personality and Social Psychology* 96: 967. [crossref]

31. Dijksterhuis A, Preston J, Wegner DM, Aarts H (2008) Effects of subliminal priming of self and God on self-attribution of authorship for events. *Journal of Experimental Social Psychology* 44: 2-9.

32. Wenke D, Fleming SM, Haggard P (2010) Subliminal priming of actions influences sense of control over effects of action. *Cognition* 115: 26-38.

33. De Houwer JD (2014) A propositional model of implicit evaluation. *Social and Personality Psychology Compass* 8: 342-53.

34. Paulhus DL (1984) Two-component models of socially desirable responding. *Journal of Personality and Social Psychology* 46: 598.

35. Paulhus DL, Reid DB (1991) Enhancement and denial in socially desirable responding. *Journal of Personality and Social Psychology* 60: 307.

36. Anderson JR (2019) The moderating role of socially desirable responding in implicit–explicit attitudes toward asylum seekers. *International Journal of Psychology* 54: 1-7. [crossref]

37. Gamberini L, Spagnolli A, Corradi N, Sartori G, Ghirardi V etal.(2014) Combining implicit and explicit techniques to reveal social desirability bias in electricity conservation self-reports. *Energy Efficiency* 7: 923-35.

38. Nosek BA (2005) Moderators of the relationship between implicit and explicit evaluation. *Journal of Experimental Psychology: General* 134: 565.

39. Riketta M (2005) Gender and socially desirable responding as moderators of the correlation between implicit and explicit self-esteem. *Current Research in Social Psychology* 11: 14-28.

40. Huntjens RJ, Rijkeboer MM, Krakau A, de Jong PJ (2014) Implicit versus explicit measures of self-concept of self-control and their differential predictive power for spontaneous trait-relevant behaviors. *Journal of Behavior Therapy and Experimental Psychiatry* 45: 1-7. [crossref]

41. Weertman A, Arntz A, de Jong PJ, Rinck M (2008) Implicit self-and other-associations in obsessive-compulsive personality disorder traits. *Cognition and Emotion* 22: 1253-75.

42. De Houwer J (1998) The semantic Simon effect. *The Quarterly Journal of Experimental Psychology Section A* 51: 683-8.

43. Hart CM, Ritchie TD, Hepper EG, Gebauer JE (2015) The balanced inventory of desirable responding short form (BIDR-16). *Sage Open* 5: 2158244015621113.

44. Jeunet C, Albert L, Argelaguet F, Lécuyer A (2018) "Do you feel in control?": towards novel approaches to characterise, manipulate and measure the sense of agency in virtual environments. *IEEE Transactions on Visualization and Computer Graphics* 24: 1486-1495. [crossref]

45. Sörqvist P (2014) On interpretation and task selection in studies on the effects of noise on cognitive performance. *Frontiers in Psychology* 5: 1249. [crossref]

46. Lieberman MD, Jarcho JM, Satpute AB (2004) Evidence-based and intuition-based self-knowledge: an FMRI study. *Journal of Personality and Social Psychology* 87: 421. [crossref]