Psicothema

Explicit and implicit assessment of gender roles

Juan Fernández, M. Ángeles Quiroga, Sergio Escorial and Jesús Privado
Universidad Complutense de Madrid

Abstract

Background: Gender roles have been assessed by explicit measures and, recently, by implicit measures. In the former case, the theoretical assumptions have been questioned by empirical results. To solve this contradiction, we carried out two concatenated studies based on a relatively well-founded theoretical and empirical approach. **Method:** The first study was designed to obtain a sample of genderized activities of the domestic sphere by means of an explicit assessment. Forty-two raters (22 women and 20 men, balanced on age, sex, and level of education) took part as raters. In the second study, an implicit assessment of gender roles was carried out, focusing on the response time given to the sample activities obtained from the first study. A total of 164 adults (90 women and 74 men, mean age = 43), with experience in living with a partner and balanced on age, sex, and level of education, participated. **Results:** Taken together, results show that explicit and implicit assessment converge. The current social reality shows that there is still no equity in some gender roles in the domestic sphere. **Conclusions:** These consistent results show considerable theoretical and empirical robustness, due to the double implicit and explicit assessment.

Keywords: Gender roles, explicit and implicit assessment, domestic sphere, gender inequalities.

Resumen

Evaluación explícita e implícita de los roles de género. **Antecedentes:** los roles de género han sido evaluados con medidas explícitas y, recientemente, implícitas. En el primer caso, los supuestos teóricos no se ajustan a los resultados empíricos. Para intentar resolver este problema hemos realizado dos estudios consecutivos con un enfoque, teórico y empírico, bien fundamentado. **Método:** el primero fue diseñado para obtener una muestra de actividades de género de la esfera doméstica mediante medidas explícitas. Los evaluadores fueron 42 adultos (22 mujeres y 20 varones, equiparados en edad, sexo y nivel educativo). En el segundo se utilizó una medida implícita, centrada en los tiempos de respuesta que las personas dan a las actividades de género, obtenidas en el primer estudio. Participaron 164 adultos (90 mujeres y 74 varones, media de edad= 43) con experiencia de vida en pareja y equiparados en edad, sexo y nivel educativo. **Resultados:** tomados conjuntamente, estos dos estudios muestran que es posible validar los datos obtenidos con una medida explícita mediante otra implícita: en la realidad social actual sigue sin haber equidad en ciertos roles de género del ámbito doméstico. **Conclusiones:** estos resultados, concordantes, manifiestan consistencia teórica y empírica, debido a la doble medida utilizada: explícita e implícita.

Palabras clave: roles de género, evaluación explícita e implícita, esfera doméstica, desigualdades de género.

Throughout the 20th century, the assessment of gender roles was mainly carried out using self-reports, at least within the field of psychology (Beere, 1990). The masculinity-femininity scales used in the first half of the past century were the most well known and most widely used instruments in the assessment of gender roles.

Towards the middle of the past century, a new approach emerged that became the theoretical foundation of the so-called *new* masculinity (M) and femininity (F) scales. This foundation would be defined as instrumentality and expressiveness or agency and communion (Helgeson, 1994; Spence & Helmreich, 1981; Stake, 1997).

The new approach led to the almost total decline of the former one (Fernández, Quiroga, Del Olmo, & Rodríguez, 2007).

However, it is not devoid of problems (Choi, Fuqua, & Newman, 2008; Fernández, 2011; Marsh & Myers, 1986).

Towards the end of the past century and the beginning of the 21st century, a new approach emerged that involves the assumption of both a new perspective (implicit mode of information processing) and new assessment instruments (basically, implicit association tests and priming measures) (Barth, 2007; Fazio & Olson, 2003; Gawronski & Payne, 2010). The new perspective focuses on implicit knowledge, and the assessment instruments consist of tasks that serve as a stimulus to record the participants' response time (RT). Responses to automated knowledge, which basically do not require introspection, as in the case of gender roles and stereotypes, are assumed to show a lower latency (Greenwald, Poehlman, Uhlmann, & Banaji, 2009; Van Well, Kolk, & Oei, 2007; White & White, 2006).

The different implicit gender role tests share a basic aspect, namely, their atheoretical character, because the emphasis is placed on the methodology rather than on a detailed analysis of the possible theories that frame the various concepts and constructs used (Cvencek, Greenwald, & Meltzoff, 2011; Greenwald & Farnham, 2000; Van Well et al., 2007).

In this paper, we assume the following theoretical assumption. Gender stereotypes and gender roles do not refer to the same entities. Gender stereotypes refer to certain belief systems about the similarities and differences between women and men, as well as the corresponding attitudes derived from these belief systems (Cuddy et al., 2009; Harris, 1995; Martin & Ruble, 2009; Tobin et al., 2010).

Gender roles refer to concrete and specific activities that men and women play more often in a society and at a given time, both in the paid workforce and in the domestic sphere. In almost all countries, men usually spend longer hours in paid work than do women, but women have to bear greater responsibility for domestic tasks (Bianchi, Milkie, Sayer, & Robinson, 2000). It is a fact that women have increased their share of paid work in almost all countries over the last half century, but men have been slow to take on a greater share of domestic responsibilities (Gregory & Milner, 2009).

The theoretical approach underlying this study assumes that: (a) There is a differentiation between gender stereotypes and gender roles, (b) gender roles are specific activities that men or women perform more frequently in a specific society and era, (c) these gender roles can be differentiated according to two different spheres: work and domestic activities, as two relatively independent fields in this historical moment (21st century), and d) there are distinct gender roles for males and females that are considered as relatively independent in this specific era.

From these assumptions, different theoretical nuclei arise: two for each conjunction of sphere and sex (Domestic Female, DF; Domestic Male, DM; Work Female, WF; and Work Male, WM) and two for the conjunction of sphere and neutral (Domestic Neutral, DN; Work Neutral, WN). Measures from the nuclei should be almost independent.

Taking into account this theoretical approach and adopting the RT methodology as the measurement level, the Gender Roles Test-36 (GRT-36; Fernández, Quiroga, Del Olmo, Aróztegui, & Martín, 2011) was elaborated. This test assumes the above-mentioned six scales or nuclei.

The present paper includes two studies. The first one was designed to carry out an explicit assessment of activities included within the domestic sphere. The main reason for analyzing only the domestic sphere is because there is a consensus in the international literature that gender inequalities are still produced much more in this area than in the labor sphere (Bianchi et al., 2000; Gregory & Milner, 2009; Wood & Eagly, 2002).

In the second study, a validation was conducted with respect to the results obtained in the first study, using the perspective of implicit knowledge and a new GRT as the instrument for implicit assessment of activities included within the domestic sphere. The essential hypothesis in this new approach is that the RT given to the items of the scales of DF and DM (genderized activities) will be significantly lower than the time given to the items of the DN scale (Greenwald et al., 2002; Greenwald et al., 2009; Van Well et al., 2007; White & White, 2006). A second hypothesis, essential to gender roles within the domestic sphere, is that most participants, both men and women, will respond according to the data obtained in the first study, that is, in the expected stereotypical way, because they are members of a society characterized by the maintenance of genderized activities within the domestic sphere (Bianchi et al., 2000; Gregory & Milner, 2009; Shelton & John, 1996). Third, if the data allow it, we will test whether non-stereotyped responses

(choosing a man for an activity typical of a woman and vice versa) will show a higher latency than stereotyped responses (Fernández et al., 2011). Fourth, it is hypothesized that women will show similar latencies to men with respect to neutral, stereotyped, and non-stereotyped stimuli because there are no sex differences in mean RT in decision-making tasks (Halpern, 2012; Jensen, 2006).

In short, the aim of the second study is to validate the results obtained in the first study, but now by means of an implicit assessment, to show that, in the domestic sphere, prevalently genderized tasks still remain, which often involve a situation of inequality for women (Wood & Eagly, 2002). We use here a new perspective (focusing on implicit knowledge) and a new instrument (focusing on the RT), assuming concordance in the results of implicit and explicit measures of gender roles instead of inconsistencies or discrepancies, as others have shown (Devos, Blanco, Rico, & Dunn, 2008; Hofman, Gawronski, Gschwender, Le, & Schmitt, 2005; Skowronski & Lawrence, 2001).

Study 1

Method

Participants

A group of 42 raters was asked to assess the adequacy of the 8 new items, as well as of the 21 items, in the domestic sphere of the GRT (6 for DM; 10 for DF; 5 for DN). To ensure anonymous participation, the raters were only asked to provide information on their sex, the age group they belonged to (greater or less than/ equal to 30 years), and their level of education (with or without a university degree). The group comprised 22 women and 20 men, mostly over 30 years old (88%). In terms of educational level, 20 participants had a university degree (48%), and 22 (52%) had none. The raters did not receive any financial compensation; they were recruited by psychology students from among their relatives and acquaintances.

No statistically significant association was found among any of the three demographic variables [Age × Sex: $\chi^2(1, N = 42) = .35$, $p = .453$; Sex × Education Level: $\chi^2(1, N = 42) = .89$, $p = .264$; Age × Education Level: $\chi^2(1, N = 42) = .35$, $p = .453$]. Thus, the sample consists of a group of Spanish adults balanced on sex, age, and educational level.

Instrument

The eight new items were selected after conducting an analysis of the domestic activities most frequently performed by women, by men, or by both. These eight (plus the 21 of GRT) were chosen as representing activities in which gender differences could exist (genderized items) or where both sexes equally performed (neutral items). These 29 items formed a new assessment booklet that each participant was asked to fill in (see Table 1). Of the 8 new activities, 6 were selected as items of the DM scale and 2 as items of the DN sphere. It was not necessary to select any items for the DF scale because it already included 10 items.

For an item to be considered as a good indicator of the set of activities for which it was designed, it had to meet the following criteria: a probability value ≥ .70 for activities relating to males and females and a probability value > .50 for neutral activities

(both sexes). Probability values refer to the probability with which participants assign the items to a predetermined target (male, female, neutral).

Procedure

Once the participant had consented to take part in the study, an individual assessment session was scheduled. To ensure complete confidentiality, ratings were performed individually. The person administering the booklet gave the instructions and clarified any questions, and afterwards remained at a discrete distance from the participant.

The instructions given to the participants were:

This study is not interested in the views of those who fill in each item, but rather what most people in our society nowadays may think about it. Examples: 1) In relation to exercise. Regardless of whether my opinion is that physical exercise does not improve health or whether I do any exercise, my answer must be based on what most people would consider. Therefore, my answer would be: I think that people believe exercise is good for their health. 2) Smoking. Even if I think smoking is not as bad as they say, or if I am a smoker, my answer would be: I think most people believe that smoking is bad for their health. Thus, please complete the following items with this guideline in mind, answering on the basis of whether people believe that this task is better done by a female or a male, or it makes no difference.

Once the assessment booklet had been completed, it was put inside an unmarked envelope, which was sealed in front of the participant, who then placed the envelope inside a portfolio containing the envelopes from other participants.

Results

Table 1 lists the percentages with which each activity was assigned to a male, a female, or both (neutral) by the group of raters. Each item is followed, in brackets, by the sphere to which it belongs (domestic) and by the sex (male or female) or category (neutral).

The results show that 26 of 29 items meet the criteria. Thus, the new domestic sphere, with its three significant nuclei, would consist of a total of 26 items, 11 belonging to the scale of DM, 10 to that of DF, and 5 to that of DN. It is important to highlight that: a) all items of the DF and DM scales adopted from the GRT-36 retained their discriminative properties; b) of the new items of the DM scale, only one did not reach the cut-off point, but no participant considered it typical of women; and c) from the neutral items of the GRT-36, only one did not appear as neutral and it was therefore removed.

Discussion

These results show that the participating raters believe that our society considers that certain activities show a bias (predetermined proportion) in favour of one sex over the other but that other activities can be equally assigned to both sexes. These data support other studies (Fernández et al., 2011) claiming that there are still clear gender inequalities within the domestic sphere

Table 1
Percentages with which Each Activity Was Assigned to Women, Men, or Both by Raters (N = 42)

GRT-26 Items	Percentage		
	Male	**Female**	**Neutral**
Hang a picture (DM*)	**83**	0	17
Change a flat tire (DM)	**100**	0	0
Walk the dog (DN)	10	11	**79**
Bring the car to the garage (DM)	**83**	0	17
Write the shopping list (DF)	0	**93**	7
Buy a present (DF)	0	**74**	26
Bring the grandfather to the doctor (DF)	0	**79**	21
Iron clothes (DF)	0	**100**	0
Play cards (mus) (DM)	**86**	0	14
Take care of the baby (DF)	0	**76**	24
Look for new apartment (DN)	2	12	**86**
Tidy the house (DF)	0	**83**	17
Fix a plug (DM)	**98**	0	2
Choose the children's clothes (DF)	0	**100**	0
Synchronize a television (DM)	**86**	0	14
Wash the floor (DF)	2	**88**	10
Pick up mail from the mailbox (DN)	16	10	**74**
Buy furniture (DN)	3	45	**52**
Prepare food (DF)	0	**90**	10
Sew the hem of a pair of trousers (DF)	0	**100**	0
Put up the beach umbrella (DM)	**90**	0	10
Fix the toilet water cistern (DM)	**95**	2	3
Mow the lawn (DM)	**93**	0	7
Change a lock (DM)	**100**	0	0
Choose a restaurant (DN)	21	19	**60**
Invest in the stock market (DM)	**83**	0	17
Removed items			
Choose a phone company (DN)	55	2	43
Buy a new car (DM; new)	62	0	38
Buy the newspaper (DN; new)	55	0	45

* DM= Domestic Male; DF= Domestic Female; DN= Domestic Neutral

in most societies (Bianchi et al., 2000; Gregory & Milner, 2009; Wood & Eagly, 2002).

The new GRT/Domestic Sphere, with its 26 items, could be used, knowing that several groups of raters (three groups in the case of the GRT-36 and one more in this study, for a total of more than 200 people) agree that these items reflect the knowledge of our society on the different gender roles.

It would be interesting to compare the cutoff points used here ($\geq$ .70 and $>$. 50) with those in other fields (advantages and disadvantages) because the decisions taken in relation with the scales of the instrument depend largely on these cutoff points. We used these cutoff points to avoid assuming that there are practically no gender roles in our Western societies at this moment (if a higher cutoff point is used) and that all the activities within the household are gendered (if a lower cutoff point is used). The obtained results showing that there are still gender roles in the domestic sphere, although the gender gap is gradually decreasing (Bianchi et al., 2000; Gregory & Milner, 2009; Shelton & John, 1996; Wood & Eagly, 2002), support the cutoff points used.

The 26 items validated in this study will be used to elaborate an implicit assessment in the second study (see Table 2).

Item number	Picture	Item	Scale/ Sphere
1		Hang a picture	DM
5		Change a flat tire	DM
8		Bring the car to the garage	DM
13		Play cards (mus)	DM
20		Fix a plug	DM
23		Synchronize a television	DM
37		Put up the beach umbrella	DM
38		Fix the WC tank	DM
39		Mow the lawn	DM
44		Change a lock	DM
52		Invest in the stock market	DM
9		Write the shopping list	DF
10		Buy a present	DF
11		Bring the grandfather to the doctor	DF
12		Iron clothes	DF
14		Take care of the baby	DF
17		Tidy the house	DF
21		Choose the children's clothes	DF
25		Wash the floor	DF
31		Prepare food	DF
36		Sew the hem of a pair of trousers	DF
7		Walk the dog	DN
16		Look for a new apartment	DN
27		Pick up mail from the mailbox	DN
30		Buy furniture	DN
49		Choose a restaurant	DN

Table 2
GRT-26 Scales: DM, DF, and DN

STUDY 2

Method

Participants

The group of participants was selected according to two main criteria: (a) married/couple life experience and (b) paid work experience. A total of 182 participants were evaluated. Of these, 5 participants were eliminated due to incomplete or erroneous data; another 13 were eliminated for being left-handed because this condition could bias the results. Participants did not receive any financial compensation. Two of the authors of this paper and a group of 5 students of psychology, especially trained for this purpose and financially compensated, were given the task of finding participants with the above-mentioned characteristics and assessing them.

The final group consisted of 164 adults, of whom 90 were female and 74 were male. Their average age was 42.73 years (*SD* = 10.34), with a range of 23 to 61 years. A total of 108 participants had a university degree, whereas 56 had none. At the time of their evaluation, 135 participants lived with their partners and their own families, 14 lived with their parents (following a divorce or for economic reasons), 9 lived alone, and 5 lived with friends.

To analyze any possible association between the demographic variables, age was dichotomized into two categories (up to 30 years and over 30 years old, to maintain correspondence with Study 1). None of the three demographic variables was found to be statistically related to any other, [Age × Sex: $\chi^2(1, N = 164) = .79, p = .491$; Sex × Education Level: $\chi^2(1, N = 164) = .18, p = .742$; Age × Education Level: $\chi^2(1, N = 164) = .53, p = .630$]. Thus, the group is non-biased in terms of socio-demographic characteristics, although it is not at all a representative sample.

Instrument

With the 26 items validated in the first study, a new instrument was elaborated. The *Gender Roles Test-26* (GRT-26) is a computerized task that can be completed online. To complete this instrument, each participant had to press either the right or the left arrow on the keyboard, depending on whether they thought that a male or a female would better perform the activity proposed by the item (see Figure 1 and Table 2). Before completing the GRT-26, the participants completed a familiarization task, which required

Example:

Buying food

Figure 1. *Sample Item from the GRT*

deciding as quickly as possible whether the image displayed is that of an animal (press the right arrow) or a plant (press the left arrow).

The GRT-26 instructions underline that the main goal of the task is to respond as quickly as possible, making the participant believe that the aim is to measure the decision-making speed and thus diverting the participant's attention from the true aim of the test (Fazio & Olson, 2003). In fact, at the end of the task, the only feedback the participant received was the RT achieved both in the preliminary familiarization task and in the male/female task.

The stimulus remains on-screen for 20000 ms. If the participant does not respond within the given time, the task moves on to the next item, and the program does not record any response except for a time of 20001 ms, which indicates that there was no response for that item.

For each item, the response is considered to be stereotyped if the person responds male to an item in the DM scale and non-stereotyped if the response is female. The same procedure applies for the DF scale. After the answers are recorded, the mean RTs for stereotyped and non-stereotyped responses are computed for each participant. In the DN scales, the mean RTs are calculated directly because no response grouping is needed.

The instrument is programmed to be applied individually online. After the GRT-26 is applied, data sets (raw and corrected) are automatically generated. These can easily be exported to data analysis programs.

To obtain the transformed variables, an SPSS syntax program has been elaborated that computes ratios, stereotyped and non-stereotyped RTs, and the within-individual variability of RTs. From hereon, researchers can compute the D scores measuring the strength of the response association (gender roles for each sphere and sex), which are necessary for individual assessment (see Fernández et al., 2011).

Procedure

Assessment assistants visited the selected workplaces, individually administered the instructions, clarified any questions, and then maintained a discrete distance while the participant performed the task. Afterwards, the collected data were inspected to detect distorted response patterns (i.e., the participants had not carefully followed the instructions) that should be discarded.

As previously mentioned, the participants provided their demographic data (see *Participants* section). Afterwards, they performed the familiarization task. Finally, the GRT-26 was administered.

Data analyses

For each of the GRT-26 scales, DF and DM, two types of scores were computed: (a) the mean RT of stereotyped responses (congruence between choice, male/female; and stereotyped activity, typically female or typically male) and (b) the mean RT of non-stereotyped responses (incongruence between choice and stereotyped activity). For the DN scale, only the mean RT was computed.

Before computing the mean RT for stereotyped and non-stereotyped responses for each participant, the RTs for each item that were less than 300 ms or greater than 6000 ms were replaced by the within-individual median (Barnett & Lewis, 1994) in the

scale to which those RTs pertain. This was done to eliminate random RTs (those not contingent with the stimulus) and those that are the result of a distraction (the participant did not follow the instructions). In some similar instruments, the criterion used is to suppress those RTs that are less than 300 ms or greater than 3000 ms (Greenwald, Nosek, & Banaji, 2003). In our case, 3000 ms is not enough because a sentence, instead of a word, is used to describe the item (see Table 2). In other studies (Fernández et al., 2011), the limit was set at 6000 ms (mean plus 2 standard deviations).

If the participant does not provide any non-stereotyped response, the non-stereotyped mean RT cannot be computed, and the program returns a missing value. In such cases, the missing values should be replaced by the mean RT of the participant to neutral stimuli in the sphere. This decision is based on the fact that the results are computed at a within-individual level; thus, it does not make any sense to replace missing values with the mean of the group in that variable or to use any other inter-individual procedure (Graham, 2009). This replacement is necessary whenever a group or an individual gender roles assessment is being followed. In this study, this procedure was not followed because one of the goals was to analyze the number of people giving both stereotyped and non-stereotyped responses for each scale and the number of people giving only stereotyped responses.

All comparisons required repeated measures techniques. For repeated measures comparisons, the alpha level was set to 0.005 after considering the family-wise error rate correction (α/n). When computing effect sizes (Cohen's *d*), dependence between means was considered and correction applied, following Morris and DeShon's (2002) formula.

Results

Regarding the first hypothesis, the obtained results show statistically significant differences both between DF and DN [M_{DF} = 1705, SD = 504; M_{DN} = 1950, SD = 571, t(163) = -8.01, p<.001, d = .597] and between DM and DN [M_{DM} = 1744, SD = 451; M_{DN} = 1950, SD = 571, t(163) = -7.27, p<.001, d = .620]. The size effects are medium to high.

For the second hypothesis, Figure 2 shows the distribution of the total number of stereotyped responses for the DF scale (Figure 1-a) and the DM scale (Figure 1-b). In both cases, the non-normality of the distributions confirms the formulated hypothesis.

For the DF scale, the standardized asymmetry is -9.89, and the standardized kurtosis is 8.81. For the DM scale, the standardized asymmetry is -8.84, and the standardized kurtosis is 8.75. For the DF scale, 21% of the group responded with 0 to 7 stereotyped responses; 33% gave 8 or 9 stereotyped answers, and 46% gave stereotyped responses on all items. For the DM scale, 21% of the group also responded with 0 to 7 stereotyped responses; 53% gave 8, 9, or 10 stereotyped responses, and 26% of the group gave stereotyped responses on all the 11 items.

With these figures, it seemed necessary to differentiate between participants who gave a low number of stereotyped responses and those who gave a high number of stereotyped answers. For this purpose, these variables were dichotomized and used as moderator variables for the analysis of the third hypothesis. In fact, due the content of the third hypothesis, only participants with a high amount of stereotyped answers were considered in this analysis. For the DF scale, 129 participants gave more than 7 responses in the direction of the stereotype, compared to 130 participants for the DM scale.
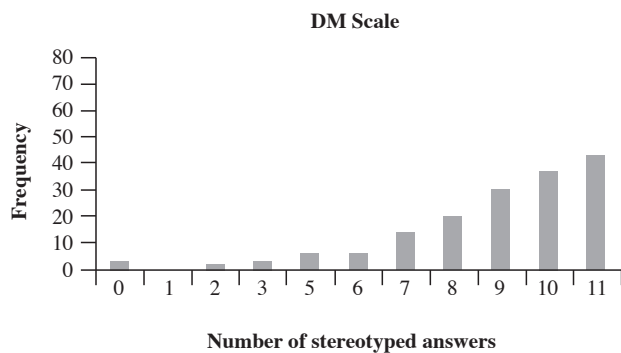
a) DF scale
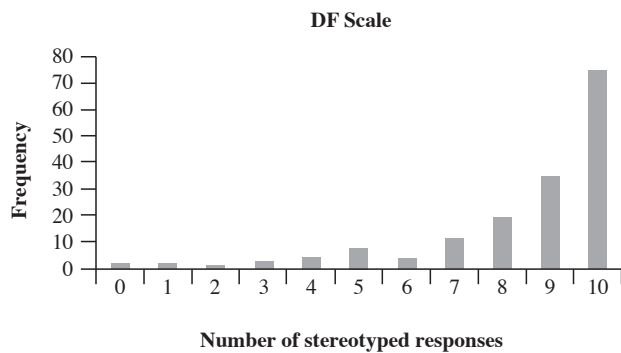
DF Scale

b) DM scale

DM Scale

Figure 2. *Distribution of total number of stereotyped responses*

For the DF scale, 54 participants were considered for the comparison between the RT when responding in the direction of the stereotype and the RT when responding in the opposite direction. The reason for this small group is that 75 participants (58% of the group with a high rate of stereotyped responses) gave all the responses in the stereotyped direction. As expected, the RT for stereotyped answers was lower [$M_{Non-Ster}$ = 1945, SD = 615; M_{Ster} = 1720, SD = 524, $t(53)$ = 2.91, p = .005, d = .399] than that for non-stereotyped responses. For the DM scale, only 87 participants were considered, because 43 (33%) gave all the responses in the stereotyped direction. Again, as expected, the RT for stereotyped answers was lower [$M_{Non-Ster}$ = 2016, SD = 722; M_{Ster} = 498, $t(86)$ = 1.74, p = .086, d = .197] than that for non-stereotyped responses. The effect size was higher for the DF than for the DM scale. Actually, the effect size for the DM scale was very low, and the differences between RTs were not statistically significant.

Table 3 displays the obtained results for the fourth hypothesis. The data show no statistically significant differences. Thus, women and men do not differ in their RTs to neutral stimuli, stereotyped responses, or non-stereotyped responses.

Discussion

It was important to verify empirically whether the responses considered by the raters of the first study as stereotyped show lower latencies in the second study than the responses given to neutral stimuli (Hofman et al., 2005). This was the case for both the DM and DF scales. This is a relatively new way of validating the results that raters produced (explicit assessment) by means of an implicit assessment: the RT to the items of the DM and DF scales elaborated with the answers the raters gave. Taking together the statistical significance and the effect size (medium/high), it seems we can state that the data obtained by an implicit assessment endorse those obtained by an explicit assessment. Both procedures are mutually reinforcing; the inconsistencies or discrepancies have disappeared (Devos et al., 2008; Hofman et al., 2005; Skowronski & Lawrence, 2001).

Within the scope of gender roles, it was important to know if the participants, members of a given society with gendered activities in the domestic sphere (according to the first study), would reflect, by means of an implicit assessment, the different types of activities that are most frequently assigned to either sex (Shelton & John, 1996). The data confirm this point. Women and men mostly chose stereotyped responses in both scales.

After these two key assumptions (one focused on the field of implicit assessment and the other in the field of gender roles), it seems appropriate to test for nuances related to both fields. Thus, as regards to the field of implicit assessments, it was noted that the data clearly support the lower latency of stereotyped responses versus non-stereotyped ones in the DF scale. We cannot state the same for the DM scale because the differences are not statistically significant, and the effect size should be considered as medium. What is the reason for this difference in the obtained results between DM and DF? The analysis of the data shows that many participants need a long time to answer the items referring to domestic activities of men; thus, the mean RT for stereotyped responses in the DM scale is quite similar to the mean RT of non-stereotyped responses in the DF scale.

Regarding the second field—that of differences/similarities in the latency of responses between women and men—, the data clearly support the hypothesis. That is, there are no differences in the RTs of males and females for stereotyped, non-stereotyped, and neutral items. These results confirm the dominance of the similarities over the differences between women and men in the performance of most of the cognitive tasks (Halpern, 2012; Jensen, 2006).

General discussion

At the present time, genderized activities still remain within the domestic sphere in Western societies. We have verified this point both with an explicit assessment methodology (first study) and with an implicit assessment one (second study). The latter is more in tune with current methodologies and is clearly *objective* (Barth, 2007; Fazio & Olson, 2003; Gawronski & Payne, 2010; Greenwald et al., 2009; Van Well et al., 2007; White & White,

Table 3
Means, standard deviations, and comparative statistics for response time to neutral, stereotyped, and non-stereotyped stimuli according to sex

	Women			Men				
	N	Mean	SD	N	Mean	SD	F	P
Neutral stimuli	90	1936	584	74	1967	561	0.12	.731
DF stereotyped answers	89	1654	509	73	1784	595	2.26	.137
DF non-stereotyped answers	39	1823	472	50	1995	621	2.07	.154
DM stereotyped answers	88	1869	483	73	1793	510	0.96	.329
DM non-stereotyped answers	73	1906	642	48	2017	757	0.75	.387

2006). The data, in turn, are part of what seems to be a clear finding of universals: there seem to be no societies in which there are no genderized differences in the domestic sphere, although (in some countries and especially throughout the second half of the 20th century) a trend of decreasing differences and rising equity within domestic sphere is evident (Bianchi et al., 2000; Fernández et al., 2011; Gregory & Milner, 2009; Wood & Eagly, 2002).

Because the software developed for this study enables online assessment and because the current data have only a very limited value (our sample was not at all representative), it seems appropriate to validate these results in various countries. The data thus obtained would enable us to have a good image of some gender roles (not all gender roles) in the domestic sphere, in terms of both their potential universality and their specificities in terms of space and time, as has been attempted with gender stereotypes (Cuddy et al., 2009). With this goal in mind, it seems useful to remove the item *Playing cards (mus)*, because a) it was formulated

in an overly specific way (it is a particular game in a particular country); b) is it an item from DM or WM?

In short, we have found genderized realities within the domestic sphere using both explicit and implicit assessments. Within this sphere, women usually perform the activities of washing, ironing, sewing, cleaning the house, and caring for children and adults, which require more time and are socially less valued (Shelton & John, 1996; Wood & Eagly, 2002). This should make us reflect about the need for actions to attain equality between women and men.

Acknowledgements

References

Barnett, V., & Lewis, T. (1994). *Outliers in statistical data* (3rd ed.). Chichester, UK: Wiley.
Barth, J.A. (Ed.).(2007). *Automatic processes in social thinking and behaviour.* New York: Psychology Press.
Beere, C.A. (1990). *Gender roles: A handbook of tests and measures.* New York: Greenwood Press.
Bianchi, S.M., Milkie, M.A., Sayer, L.C., & Robinson, J.P. (2000). Is anyone doing housework? Trends in the gender division of household labor. *Social Forces, 79,* 191-228.
Choi, N., Fuqua, D.R., & Newman, J.L. (2008). The Bem Sex-Role Inventory: Continuing theoretical problems. *Educational and Psychological Measurement, 68,* 881-900.
Cuddy, A.J.C., Fiske, S.T., Kwan, V.S.Y., Glick, P., Demoulin, S., Leyens, J.P., …, Ziegler, R. (2009). Stereotype content model across cultures: Toward universal similarities and some differences. *British Journal of Social Psychology, 48,* 1-33.
Cvencek, D., Greenwald, A.G., & Meltzoff, A.N. (2011). Measuring implicit attitudes of 4-year-olds: The Preschool Implicit Association Test. *Journal of Experimental Child Psychology, 109,* 187-200.
Devos, T., Blanco, K., Rico, F., & Dunn, R. (2008). The role of parenthood and college education in the self-concept of college students: Explicit and implicit assessment of gedered aspirations. *Sex Roles, 59,* 214-228.
Fazio, R.H., & Olson, M.A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology, 54,* 297-327.
Fernández, J. (2011). Un siglo de investigaciones sobre masculinidad y feminidad: una revisión crítica [A century of research on masculinity and femininity: A critical review]. *Psicothema, 23,* 167-172.
Fernandez, J., Quiroga, M.A., Del Olmo, I., & Rodríguez, A. (2007). Escalas de masculinidad y feminidad: estado actual de la cuestion [Masculinity and femininity scales: Current state of the art]. *Psicothema, 19,* 357-365.
Fernández, J., Quiroga, M.A., Del Olmo, I., Aróztegui, J., & Martín, A. (2011). Objective assessment of gender roles: Gender Roles Test (GRT-36). *The Spanish Journal of Psychology, 14,* 899-911.
Gawronski, B., & Payne, B.K. (Eds.) (2010). *Handbook of implicit social cognition: Measurement, theory, and applications.* New York: Guilford.
Graham, J.W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology, 60,* 549-576.
Greenwald, A.G., Bahaji, M.R., Rudman, L.A., Farnham, S.D., Nosek, B.A., & Mellott, D.S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review, 109,* 3-25.

Greenwald, A., & Farnham, S.D. (2000). Using the Implicit Association Test to measure self-esteem and self-concept. *Journal of Personality and Social Psychology, 79,* 1022-1038.
Greenwald, A.G., Nosek, B.A., & Banaji, M.R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85,* 197-216.
Greenwald, A.G., Poehlman, T.A., Uhlmann, E.L., & Banaji, M.R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology, 97,* 17-41.
Gregory, A., & Milner, S. (2009). Editorial: work-life balance: A matter of choice? *Gender, Work and Organization, 16,* 1-13.
Halpern, D. (2012). *Sex differences in cognitive abilities* (4th ed.). New York: Psychological Press.
Harris, J.R. (1995). Where is the child's environment? A group socialization theory of development. *Psychological Review, 102,* 458-489.
Helgeson, V. (1994). Relation of agency and communion to well-being: Evidence and potential explanations. *Psychological Bulletin, 116,* 412-428.
Hofman, W., Gawronski, B., Gschwender, T., Le, H., & Schmitt, M. (2005). A meta-analysis of the correlation between the Implicit Association Test and explicit self-report measures. *Personality and Social Psychology Bulletin, 31,* 1369-1385.
Jensen, A.R. (2006). *Clocking the mind: Mental chronometry and individual differences.* Oxford, UK: Elsevier.
Marsh, H.W., & Myers, M.R. (1986). Masculinity, femininity and androgyny: A methodological and theoretical critique. *Sex Roles, 14,* 397-430.
Martin, C.L., & Ruble, D.N. (2009). Patterns of gender development. *Annual Review of Psychology, 61,* 353-381.
Morris, S.B., & DeShon, R.P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods, 7,* 105-125.
Shelton, B.A., & John, D. (1996). The division of household labor. *Annual Review of Sociology, 22,* 299-322.
Skowronski, J.J., & Lawrence, M.A. (2001). A comparative study of the implicit and explicit gender attitude of children and college students. *Psychology of Women Quarterly, 25,* 155-165.
Spence, J.T., & Helmreich, R.L. (1981). Masculine instrumentality and feminine expressiveness: Their relationships with sex roles attitudes and behaviours. *Psychology of Women Quarterly, 5,* 147-163.
Stake, J.E. (1997). Integrating expressiveness and instrumentality in real life settings: A new perspective on the benefits of androgyny. *Sex Roles, 37,* 541-564.

Tobin, D.D., Menon, M., Menon, M., Spatta, B.C., Hodges, E.V.E., & Perry, D.G. (2010). The intrapsychics of gender: A model of self-socialization. *Psychological Review, 117,* 601-622.

Van Well, S., Kolk, A.M., & Oei, N.Y.L. (2007). Direct and indirect assessment of gender role identification. *Sex Roles, 56,* 617-628.

White, M.J., & White, G.B. (2006). Implicit and explicit occupational gender stereotypes. *Sex Roles, 55,* 259-266.

Wood, W., & Eagly, A.H. (2002). A cross-cultural analysis of the behaviour of women and men: Implications for the origins of sex differences. *Psychological Bulletin, 128,* 699-727.