Documento de Trabajo/
Working Paper

IESA 02-04

Effects of Gender and Age
on Retrospective Time Judgements

Pablo Brañas-Garza
Departamento de Economía Aplicada, Universidad de Jaén, Spain. E-mail:
pbg@ujaen.es

Antonio J. Morales
Departamento de Teoría e Historia Económica, Universidad de Málaga, Spain. E-mail:
amorales@uma.es

Rafael Serrano-del-Rosal
Instituto de Estudios Sociales Avanzados, Consejo Superior de Investigaciones Científicas, IESA/CSIC. Spain. E-mail: r.serrano@iesaa.csic.es
Abstract

This study provides additional empirical evidence to the research concerning the effects of gender and age on retrospective time judgements using data obtained from a Spanish database with more than 40,000 individual observations on time estimations. Statistical analyses were performed using the Levene F-test and the t-test of variances and means, respectively. The most important results of the study are as follows: i) differences in time estimation in relation to gender largely depend on the age groups analysed, with greater differences observed in the younger age group; and ii) a monotonous relationship does not exist between age and time estimations when age is a constituent factor in time judgements.

Key words: continuous time, round numbers, retrospective, prominence, gender, ageing.

1 Introduction

The notion that time is a continuous variable, and as such, is possible to divide into infinite fractions is a widely-accepted fact that no longer comes as a surprise. However, research into the human process of estimation suggests that individuals have enormous difficulties in perceiving time as a continuum.

A great many researchers across disciplines have dealt with the issue of time measurement. Psychologists, sociologists, philosophers, mathematicians and physicists, among many others, have been concerned with understanding how individuals calculate different variables.

Individual time judgement is a complex issue as individuals require a large capacity for estimation in order to perform prospective time judgements. From a retrospective viewpoint, the problem is more pronounced given that time estimation tasks require not only an exercise in computation, but memory as well. Thus a question such as “How long have you been waiting for the bus?” demands that at least three tasks be performed: calculating arrival time or the “starting point” (see A in diagram 1), calculating the time when the bus was caught or the “ending point” (see B in diagram 1), and the difference between both magnitudes.

In addition to the problems involved in guessing the starting (A) and ending (B) points, the third task (difference) can present enormous difficulties. Bounded rationality individuals may be incapable of performing these calculation tasks and will “round off”, resulting in “systematic” computational errors (see Brañas-Garza and Morales, 2003).

However, these two memory tasks are extremely complex in and of themselves as they require the sequencing of all actions performed during a given time interval (see Diagram 1). These tasks not only require the use of memory, but also a certain amount of skill in assigning the time required for each action. Clearly, there are also computational difficulties involved in doing this.
The field of psychology has provided us with enormous insight into time measurement under laboratory conditions from both prospective and retrospective approaches. From a prospective approach, the basic problems of time estimation without the use of memory have been examined through experiments dealing with the calculation of 10-second units of time such as: “When I tell you, count off 10 seconds. When 10 seconds have elapsed, raise your hand” (see, for example, Espinosa-Fernández et al., 2003). Nevertheless, while these studies analyse an essential aspect of the issue, they do not examine the effects of guessing tasks (guessing the starting point and the ending point).

In their comparative review of retrospective and prospective duration judgements, Block and Zackay (1997) highlighted three essential ideas: i) from a prospective approach, time judgements are generally longer; ii) individuals tend to underestimate time when both approaches are used; and iii) time judgements are more accurate when a prospective approach is used.

On the whole, psychological research has found that gender differences do indeed exist in time judgements. In general, men are more accurate and homogeneous in their estimations, regardless of the method used, be it verbal estimation, repeated production, reproduction or comparison (Block et al., 2000; Bell, 1972; Eisler and Eisler, 1992; Rammsayer and Lustnauer, 1989).

Gender, however, is not the only factor to be taken into account when examining time estimation. In fact, many experiments conclude that, with age, individuals increasingly underestimate the standard (see Fraisse, 1984). On the other hand, Block et al. (1997) observed that the effects of age depend on the method employed. Specifically, no differences were observed when the reproduction method was used.

Nevertheless, few experiments have examined both effects jointly. Despite the potential interest of this type of study, given that it could provide additional evidence for experiments examining these effects independently, little research has been done along these lines (Block et al., 2000; Espinosa et al., 2003).

The aim of this paper is, therefore, to offer additional retrospective evidence concerning the effects of gender and age (and gender-age) on time judgements. To this end, a database with more than 40,000 observations on time estimation has been used. The database was obtained from a survey in which individuals were asked how long they had waited to enter the doctor’s office. Respondents were asked to give a numerical value (in minutes) to this time judgement (hereafter TJ).

Two-thirds of the observations involved a time task in which the starting point (A in diagram 1) was known. In other words, respondents knew the time of their appointment and simply had to calculate the difference between the time that they should have entered and the time that the actually entered, that is, the time elapsed between the two. In this case, the time judgement was limited to guessing the actual moment of entry to the doctor’s office (note that B was estimated—that is, β—as it was not known) and calculating the difference between both magnitudes (TJ = β – A).

1 Note that when TJ = 0 → β = B = A, that is, the entry time is the same as the estimated time, which is, in turn, the same as the ex-post entry time. Therefore, there is no waiting time.
The remaining observations correspond to individuals who went to the doctor without an appointment and therefore did not have an exogenously fixed reference point or starting point (see Diagram 1). Given that the value of A was unknown, individuals were required to estimate it, \( \hat{a} \). In the same manner, they were required to estimate the moment of entry into the doctor’s office (\( \beta \)). Finally, individuals were also required to calculate the time judgement: \( TJ = \beta - \hat{a} \).

The most significant results of the study are as follows:

1. Sex differences in time estimations wholly depend on the age group analysed, with greater differences observed in the youngest group.

2. Although age is a constituent factor in time judgements, a monotonous relationship does not exist between age and time estimation.

3. The existence of a reference point, i.e. the fact that the starting point is known (A) or estimated (\( \hat{a} \)), seems to be of little importance.

Following the introduction, the study is organised into four sections. Section two examines the difficulties involved in measuring time, while the database is described in section three. Finally, the results are discussed in section four and conclusions are reached in section five.

2 Measuring time

Young and Ziman (1970) found that “clock” or analogical time becomes continuous as a result of successive divisions. However, due to its particular nature, an analogical clock fractionally differentiates time and only through the continuous fragmentation of periods of time is it possible to perceive time as a continuum.

Unfortunately, fractioning time is not a simple task, especially when taking into account the bounded rationality of individuals. When individuals are asked to perform tasks involving computations in the continuum, they tend to simplify these tasks by using strategies such as “round numbers” or “prominence numbers”. The former strategy refers to the fact that individuals tend to round off magnitudes when performing operations. For example, the conversion from euros to dollars is usually done on a one to one basis although the actual exchange rate varies daily. This “rule” of equivalence between currencies is not due to a lack of information (current exchange rates can be easily found in any newspaper) but rather to the convenience of performing simple and straightforward calculation tasks.

The theory of prominence numbers is somewhat more complex. Based on a study by Albers & Albers (1983), Selten (1994) found that individuals encounter difficulties when performing numerical tasks in intervals of less than 10 units, while intervals of 2.5 are virtually impossible. Faced with this difficulty, individuals seek “convenient” intervals to perform these tasks, which then becomes the determinant interval of their calculations. For example, suppose an individual is asked to calculate how tall Reinhard Selten is. Firstly, the individual must set an interval, for example between 100 and 200 cm. Secondly, he will fix
a middle point, say 150 cm. If there is no reason to think that Reinhard Selten is neither shorter nor taller than this middle point, the process will conclude. However, if the individual thinks, for example, that Reinhard Selten is taller, he will go back and fix a new interval, say between 150 and 200 cm. A new middle point of 175 cm will then be set and the individual will again decide if Reinhard Selten is shorter or taller. If he does not think so, he will stop. However, if the individual thinks Reinhard Selten is taller, he will recalculate the interval once again, setting it at 175-200 cm, and so on until there is no longer any reason to carry on.

Note that the new interval (175-200) is less “manageable” as it contains decimal numbers, making the task more difficult. However, there is an additional problem. Could we have arrived at the same point if the initial interval was 160-190? The answer is probably not.

This example illustrates the complexity of measuring magnitudes. Note that calculating time is not much different from calculating this Nobel prizewinner’s height. Let us return to the example of the bus. If an individual wants to know how long he has been waiting for the bus, he must first know when he arrived at the bus stop (starting point or \( A \) in diagram 1) and when he boarded (ending point or \( B \)). Only when both times have been estimated (i.e. \( a \) and \( b \) ), will it be possible to calculate the time elapsed between both points.

When calculating the starting point (\( a \)), we encounter a similar problem to that involved in estimating Reinhard Selten’s height. When did the individual arrive at the bus stop? Suppose the interval is set between 4 p.m. and 5 p.m., the middle point would then be 4:30. If there is no reason for the individual to think that he arrived before or after this time, he will stop. But, if he thinks that he arrived before 4:30, he will go back and recalculate the interval, and so on. The same problem occurs when estimating the ending point (\( b \)). However, one last calculation remains to be done: the difference between both points: \( TJ = b - a \).

All of these operations or timing tasks make time estimation particularly complex. Given this complexity, individuals tend to use simple rules to facilitate their calculations. By using a methodology similar to that of prominence numbers, Brañas-Garza, Morales & Serrano (2003) have demonstrated that most respondents in the database use multiples of 5 when estimating time.

3 Database and preliminary considerations

3.1 Database

The database used in this study, “Improving Patient Satisfaction”, was compiled by the Institute of Advanced Social Studies (CSIC) in Spain with funding from the Department of Health of the Andalusian Regional Government.

The sample was randomly selected from a representative population in the region of Andalusia. Between May 2000 and June 2002, 46,757 users of the region’s public health care service were personally interviewed at the entrance to health centres. Each

\footnote{Note that when the individual knows ex-ante the initial reference point (starting point or \( A \)) the problem is simplified. In order to perform time judgements, two possible strategies can be used: i) estimating the ending point (\( b \)) and the difference between this point and \( A \) (which is known), \( TJ = b - A \); and ii) using a shortcut: the subjective evaluation of the time elapsed without calculating the above difference.}
The questionnaire consisted of 50 questions regarding the quality of service, user satisfaction with the treatment received by health care professionals, hospital facilities, etc.

Prior to analysing the data, it should be stated that there are two ways of arranging to see a general practitioner or pediatrician: by appointment or by number.

- The appointment system: patients previously request an appointment for a set time and date.
- The number system: patients arrive at the health centre and “wait in line” until they are attended to. In this case, entry is strictly on a first come, first serve basis.

Henceforth, the first group will be referred to as the ‘appointment’ group, while the second group will be referred to as the ‘number’ group. 66.4% of the sample population in the database belonged to the ‘appointment’ group, while the rest had not made an appointment.

3.2 Variables

Question P40 of the survey asked individuals who had made an appointment:

“How long did you have to wait from the given appointment time until you actually entered the doctor’s office? ………………… minutes.

For individuals who had not made an appointment question P43 asked:

“How long did you have to wait from the time you arrived at the centre until you entered the doctor’s office? ………………… minutes.

Given the difficulties associated with making time judgements ($TJ$) as seen above, we found that:

i) Respondents with an appointment (question P40) knew their appointment time or the time that they were supposed to enter. Therefore, the starting point ($A$ in diagram 1) was known. In order to estimate their waiting time, respondents were required to calculate the moment that they actually entered ($\beta$ must be estimated by guessing) and subsequently compute the waiting time by calculating the difference between the ex-ante and ex-post entry time ($TJ = \beta - A$).

ii) Respondents without an appointment (question P43) faced a more complex problem. Not only did they have to calculate $\beta$, but also perform a similar task to estimate $\hat{\beta}$. After estimating both times, the difference was then calculated: $TJ = \beta - \hat{\beta}$.

---

3 This system is gradually being replaced by the appointment system and only occurs in health centres that have not yet established an appointment system.
The gender variable was included in question CL1. Special importance was not given to obtaining an equal number of observations from men and women as the survey was designed with the objective of reproducing real user frequency at the health centres. However, little disparity was observed as 61.1% of the observations corresponded to women (see table 1).

The CL2 question corresponded to the age of the respondents. As in the question above, little importance was given to equal proportions across ages. Nevertheless, all age groups were well represented (see table 2).

Finally, a total of 46,757 individuals were interviewed (including respondents with an appointment and those with only a number). These respondents gave waiting times in time intervals of time, \( t \in [0.300] \) minutes. Given that the majority of the population is to the left of the interval, we decided to limit the sample population to respondents who gave times of less than or equal to 100 minutes, \( t \in [0.100] \), leaving us with a total of 45,697 observations.

3.3 Additional questions

The research developed throughout the study is based fundamentally on the following two points:

1. The information collected from question 40 and question 43 refers to the time that the respondents said had elapsed. In other words, we do not know the real time, only the estimations given by the respondents themselves.

2. The objective of the study was to compare the population distributions which resulted from controlling the characteristic to be analysed (for example, gender). Using parametric tests, we were able to determine if these characteristics affect time estimation and whether or not the subjects came from the same population. That is, we analysed if this characteristic was significant.

This method might have been questionable had the sample been small, for example 100 subjects. Had this been the case, individuals with different characteristics would have waited for different periods of time. For example, all the women in the sample (50 if they comprised half the sample) would have waited longer than the men. However, given that the sample was randomly selected from a representative population of more than 40,000 observations, this situation is virtually impossible in statistical terms.

Before examining the results of the study, the following issues should be considered:

3. The great majority of respondents (43,740 or 95.7%) used round numbers when estimating their waiting time, i.e. numbers ending in 0 or 5.

4. 40,845 respondents used clock numbers (i.e. 0, 5, 10, 15, 30 and 60 minutes). In other words, 90% of the respondents used analogical (discreet) time schemes (see Brañas-Garza et al., 2003 for a detailed analysis of this type of measurement).
In order to achieve the greatest possible homogeneity, only the last group, representing 90% of the sample, will be examined below. Thus, statistical problems such as empty cells will be avoided and the study will be more general. Therefore, the sample will be limited to 40,917 observations.

5. Finally, prior to the analysis that follows, we will examine both samples (appointment and number) to determine if they do, in fact, have distinct distributions. The subjects in the ‘appointment’ sample waited, on average, for 19.43 minutes (st. dev. = 17.12), while those in the ‘number’ sample waited an average of 25.14 minutes (st. dev. = 20.34). In order to compare variances, a Levene test was used ($F = 670.33$), indicating that the variances were different ($p – value = 0.00$). The t-test of means also verified that the distributions were distinct ($t = -28.15; p – value = 0.00$). Hence, the observations did not come from the same population. This result is perfectly plausible if we consider that the use of reference points simplifies time estimation and leads to greater accuracy.

In the section that follows, the effects of gender and age on time estimation will be examined with relation to individuals with and without a reference point.

4 Results

4.1 Gender differences

The meta-analytic review by Block et al. (2000) on retrospective time studies shows that sex differences do indeed exist. Women, in particular, were observed to overestimate time to a greater degree than men (see Block et al.: p.1340). Using a prospective analysis, Espinosa-Fernández et al. (2003) obtained the same gender bias in 5 minutes estimations.

In this section the gender effect will be compared in the two samples: appointment ($A$ is known) and numbers ($\hat{a}$ is estimated). The descriptive and statistical analysis of the gender effect for both samples is shown in Table 1.

4.1.1 Appointment sample

The Levene F-test rejected the equality of variance, indicating that there are, in fact, differences between men and women in accurately estimating time. That is, men are somewhat more skilful in performing timing tasks. However, the difference in means is only comparable for a value of $\alpha = 10\%$. This would seem to suggest that, overall, the sex effect is not highly pronounced. Therefore, these results are not very conclusive.

[Insert table 1 about here]

However, given the large size of the sample, we were able to make some alternative sub-samples in order to determine if the effect continued to be ambiguous, or if it was possible to observe clearer differences.
• If the sample size is reduced to the response interval time, \( t \in [0.60] \) minutes, the Levene test weakly rejects the equality of variances \( (F = 3.25, p-value = 0.07) \), while the t-test strongly rejects the equality of means \( (t = 2.77, p-value = 0.00) \).

• When examining the first 30-minute interval, \( t \in [0.30] \) minutes, the Levene test weakly rejects the equality of variances \( (F = 2.82, p-value = 0.09) \), while the t-test strongly rejects the equality of means \( (t = 5.69, p-value = 0.00) \).

• Finally, if the interval is further reduced, \( t \in [5.30] \) minutes, the results are identical: the Levene test weakly rejects the equality of variances \( (F = 3.80, p-value = 0.05) \) and the t-test strongly rejects the equality of means \( (t = 3.98, p-value = 0.00) \).

To summarize, only when the time intervals are very large do men and women estimate in a similar way and no marked differences are observed. However, when the time intervals are less than or equal to an hour, gender differences are observed. Interestingly, these differences are more significant in terms of the mean rather than the variance, indicating that there is a sex error or gender bias.

In contrast to the results obtained by the prospective literature, in our study, women systematically remain below the average, that is, they underestimate time more than men. In terms of variance, the differences are not as evident as the sample dispersion is similar for both men and women.

**Result 1.** *When individuals have a reference point (A) and the time interval is less than or equal to an hour, a gender bias occurs and women underestimate time to a greater degree than men.*

4.1.2. Number sample

As shown in Table 1, the Levene test rejects the equality of variances for a value of \( \alpha = 5\% \), while the t-test only rejects the equality of means at 10\%. Thus, the differences between men and women when estimating time without a reference point are not as pronounced. However, as seen above, the use of smaller time intervals leads to distinct results. Let us now determine whether a reduction in the interval is significant in the group of individuals without a reference point.

• If the interval is reduced, \( t \in [0.60] \) minutes, the Levene test strongly rejects the equality of variances \( (F = 7.36, p-value = 0.00) \), while the t-test only weakly rejects the equality of means \( (t = -1.67, p-value = 0.09) \).

• With a reduced time interval, \( t \in [0.30] \) minutes, the results are clearer. The Levene test weakly rejects the equality of variances \( (F = 3.76, p-value = 0.05) \) and the t-test strongly accepts the equality of means \( (t = -0.02, p-value = 0.97) \).

• When the time interval is further reduced, \( t \in [5.30] \) minutes, the results are verified. The equality of variances continues to be weak \( (F = 3.34, p-value = 0.06) \), while the equality of means no longer exists \( (t = -0.48, p-value = 0.62) \).
In short, for time intervals of less than an hour, no gender differences occur in time estimation. This absence of the gender effect is due to the lack of a reference point.

**Result 2.** *When men and women do not have a reference point to estimate time, there is no gender bias.*

**Result 3.** *When estimating time intervals greater than an hour, there is no gender effect and the reference point is irrelevant.*

### 4.2 Ageing differences

Evidence suggests that age affects time estimations. Specifically, individuals underestimate the standard (they lose estimation skills) with age (see Fraisse, 1984). However, Block et al. (1998) observed that the age effect wholly depends on the method employed. For example, when a verbal method of estimation is used, elderly individuals overestimate, while the opposite occurs when using the production method.

According to Fraisse (1984), the tendency of elderly individuals to underestimate is related to a loss of the neuronal capacity needed to perform these tasks.

However, learning theory suggests the contrary: their own experience. Thus, older individuals may possess certain skill in implementing their own rules of estimation; rules that have been perfected throughout their lives.

The intervals proposed by Espinosa-Fernández et al. (2003) were used in this study to classify individuals in relation to age. The descriptive statistics are shown in Table 2A, while a summary of the tests is given in Table 2B.

#### 4.2.1. Appointment sample

Table 2A (lines 1 to 8) shows the average values by age group for individuals using a reference point (the starting point \( A \) is known) to estimate waiting time. Figure 1 graphically illustrates these averages (see the line with diamonds). As can be seen, given that there are several peaks, the average does not decrease in a clear manner. Thus there is no decreasing (monotonous) trend between age and time estimation.

[Insert table 2A about here]

Both young individuals between the ages of 10 and 30 and the elderly (older than 81) overestimated waiting time. However, individuals between the age of 41 and 50 also overestimated this time interval. It is interesting to note that the lower values correspond to individuals between the age of 31 and 40, whereas the remaining groups have intermediate values (see Figure 1).

[Insert Figure 1 about here]

---

4 There are two differences between the classification scheme proposed by Espinosa-Fernández et al. and the one used in this study: 1) the 8-10 year old and 11-20 year old groups were merged into one sample group as there were very few observations available for the first group; 2) due to the availability of data, a new category was added for >81 year -old individuals.
Given that the differences are not as pronounced, statistical tests were used to determine if the distribution comparisons could provide additional evidence. The results are summarised in Table 2B (Sample 1). The rows show the comparisons between successive categories, while columns 2 and 3 analyse the Levene test and the t-test (where NR accepts the null hypothesis, R is the full rejection, R* significant for $\alpha = 5\%$ and R** for $\alpha = 10\%$).

We will now centre our analysis on a comparison of the averages. The following ideas can be extrapolated from table 2B.

- Significant differences (in the averages) were observed between the 10-20 year-old group and the 21-30 year-old group.
- Strong significant differences were seen between individuals aged 21-30 and 31-40.
- Similar differences were observed between individuals aged 31-40 and 41-50.
- No significant differences were found from the age of 50 and above.

To sum up, although there are certain differences between age groups, no specific pattern regarding time estimation was observed. Therefore, this finding cannot be generalised as a monotonous trend associated with age.

Result 4. *There is no monotonous (increasing or decreasing) trend between time estimation with a reference point and age.*

This result is, on the whole, in accordance with the findings by Block et al (1998).

4.2.2. Number sample

Table 2A (lines 9 to 18), Table 2B (sample 2) and Figure 1 (line with squares) above illustrate the behaviour of individuals grouped by age with no reference point (starting point $\hat{a}$ is unknown). Again, a defined pattern associated with time estimation and age was not observed.

However, it is surprising to note the same groups that overestimated time in the previous sample, do so again in this second sample, albeit without a reference point.

Although the magnitudes are not identical, the age groups wholly coincide when over or underestimating time. The results obtained are therefore similar: there is no pattern of association between age and time estimation without a reference point.

The parametric analysis used to study differences between distributions shed some additional light on this question (see table 2B, sample 2 for a summary of the results).

- No differences (in averages) were found between individuals aged 10-20 and 21-30.
- Weak differences were observed between individuals aged 21-30 and 31-40.
- Significant differences were found between individuals aged 31-40 and 41-50.
• Significant differences were found between individuals aged 41-50 and 51-60.

• No differences were observed between individuals aged 60 and above.

When comparing these results with those of the above section, some differences were observed as young individuals do not behave in a similar manner. However, these differences do not provide much insight into the relationship between age and time estimation.

**Result 5.** *There is no monotonous trend (increasing or decreasing) between time estimation without a reference point and age.*

Finally, all the tests in Table 2B (samples 1 and 2) and Figure 1 show that there exist certain differences between the young, adults and the elderly when estimating time. Taking into account that this relationship is not monotonous, it can be said that for some reason (we do not know why) certain age groups estimate time in a different manner. The tendency of the young, adults and the elderly to overestimate time (or to estimate it accurately as in sample 1) could be explained by the greater neuronal capacity of younger individuals, by the adults’ occupation (and, therefore, the opportunity cost of their time) and finally, by the accuracy of the rules used by the elderly which have been perfected over time. Interestingly, these effects are independent of the existence of reference point.

**Result 6.** *There is a factor associated with age, regardless of the existence of a reference point, which affects time estimation. This factor neither increases nor decreases with age in a linear manner.*

### 4.3 Age and sex

The prospective study by Espinosa-Fernández *et al.* (2003) jointly analyses the effect of age and sex on time estimation skills. However, little research along these lines has been done from a retrospective approach.

Given that the database used here permits this type of comparison, we will analyse sex and age jointly in this last sub-section. Using the results from previous sections, we will limit our study to three age groups: under 31, 31-60 years of age and 61 and above. The results of the statistical analysis are shown in Tables 3A and 3B.

#### 4.3.1 Appointment sample

The effect of age in the group of women is analysed in Table 3A (columns 2 and 3 for sample 1), specifically, young women, adults and the elderly. The same analysis is shown in Table 3A for men.

• The results of the first comparison, young women vs. adult/elderly women, are evident: significant differences occur in time estimation between these age groups.

• Identical results were found when comparing young men and adult/elderly men.
• However, these differences were not observed when comparing adult men or adult women with elderly individuals.

**Result 7.** An age effect was observed in time estimation with a reference point \( (A) \) when comparing young individuals \((\text{age}<31)\) with adults \((\text{age}\geq31)\) and elderly age group \((\text{age}\geq61)\). These differences were independent of sex.

[Insert table 3A about here]

As can be seen, this result is analogous to that shown graphically in Figure 1. Thus very high values were observed for young individuals, while average values (either above or below the mean) were observed for the remaining age groups. The population of these age groups (the young, adults and the elderly) is classified in Table 3B and gender bias is analysed. The following ideas can be summarized:

• In the group of young individuals (under 31 years of age), significant differences were observed between men and women (for a value of \( \alpha \geq 5\% \)).

• In the adult group \((30<\text{age}<61)\) gender biases in time estimation were not observed.

• Finally, no gender bias was found in the elderly age group.

**Result 8.** Gender biases in time estimation with a reference point \( (A) \) occur at a young age. However, from the age of thirty and above, this gender effect does not exist.

[Insert table 3B about here]

### 4.3.2. Number sample

The population sub-sample without a starting point \( (\hat{a} \text{ is estimated}) \) does not appear to differ greatly from that of the previous section. Table 3A (columns 4 and 5) shows a comparison by age for males followed by females.

• No significant differences were found between young men and adult males in time estimation, or between adult males and elderly males.

• There are 5% differences between young women and adult women when estimating time intervals.

• Strong differences between young and elderly women are observed.

• However, no differences were observed between adult women and elderly women.

In short, differences in time estimation (without a reference point) across ages were only observed between young women and adult/elderly women. No age effects were observed in males.
Result 9. An age effect in time estimations without a reference point was only observed between young women and adult/elderly women. Significant results were not obtained in the rest of the cases.

Gender biases are analysed in table 2B (columns 4 and 5). The comparison between males and females in the three age groups indicates that:

- A gender bias exists in the group of youngest individuals (under 31)\(^5\).
- No differences were observed between adult males and adult females (30<age<61).
- No differences in means were found between males and females in the elderly group.

These results are practically identical to those analysed in the previous section (see result 8). A gender bias was only observed in the group of young individuals, while no differences were found between men and women in the other age groups.

Result 10. Gender bias in time estimation occurs at a young age, regardless of the existence of a reference point.

This last finding contradicts the evidence proposed in the literature. Specifically, Espinosa-Fernández et al. (2003), citing Block et al. (2000), suggests that:

1. Time estimation varies with age. However, results 7 and 9 of our study indicate that this is true only when young individuals are compared with adults, but not in the rest of the cases.

2. Gender bias is completely absent in adolescents and young adults, while the opposite is true in the elderly. However, results 8 and 10 of our study show that gender bias occurs precisely in the youngest age group. In other words, gender differences in time estimation only occur in individuals under 30 years of age.

5 Conclusions

The objective of this study was to provide additional empirical evidence to the study of gender and age in time estimation from a retrospective paradigm. With this aim, a Spanish database with more than 40,000 individual observations on time judgements was used. Two-thirds of the observations corresponded to respondents who used a reference point (starting point) when estimating time, whereas the rest performed estimations with no reference point.

Note that this result does not contradict result 2, but clarifies it. For the whole population – without a breakdown by age - the test did not show a gender bias. However, when the sample is limited to a smaller number of observations (only young individuals), differences are observed.
The sample was broken down by group according to the effect to be studied. The resulting distributions were then analysed statistically and compared using the Levene F-test and the t-test to determine the variances and means, respectively. The most significant results of the study results are as follows:

1. Differences in time estimation related to age are relative. The most significant differences were observed when comparing young individuals and those older than 30 years of age.

2. A monotonous relationship does not exist between age and time estimation.

3. Gender bias is more pronounced in young individuals than in adults, regardless of the existence of a reference point.

In conclusion, our findings suggest that individuals learn to estimate time during their youth.
REFERENCES


Reinhard Selten, "Descriptive Approaches to Cooperation", in Cooperation: Game-Theoretic Approaches (S.Hart and A. Mas-Colell Eds.), Springer NATO, 1994.

DIAGRAM 1: RETROSPECTIVE TIME JUDGEMENT

Time Judgement = B - A
FIGURE 1: AGEING EFFECT (SAMPLE 1 & 2)
### TABLE 1: GENDER DIFFERENCES

#### SAMPLE 1: APPOINTMENT

<table>
<thead>
<tr>
<th></th>
<th># SUBJECTS</th>
<th>AVERAGE</th>
<th>ST.DEV.</th>
</tr>
</thead>
<tbody>
<tr>
<td>MALE</td>
<td>10,270</td>
<td>19.66</td>
<td>16.76</td>
</tr>
<tr>
<td>FEMALE</td>
<td>16,943</td>
<td>19.29</td>
<td>17.33</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TESTS</th>
<th>STAT.</th>
<th>P - VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>VARIANCE</td>
<td>F - LEVENE</td>
<td>8.63</td>
</tr>
<tr>
<td>AVERAGE</td>
<td>T- TEST</td>
<td>1.72</td>
</tr>
</tbody>
</table>

#### SAMPLE 2: NUMBER

<table>
<thead>
<tr>
<th></th>
<th># SUBJECTS</th>
<th>AVERAGE</th>
<th>ST.DEV.</th>
</tr>
</thead>
<tbody>
<tr>
<td>MALE</td>
<td>5,628</td>
<td>24.80</td>
<td>20.11</td>
</tr>
<tr>
<td>FEMALE</td>
<td>8,004</td>
<td>25.38</td>
<td>10.49</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TESTS</th>
<th>STAT.</th>
<th>P - VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>VARIANCE</td>
<td>F – LEVENE</td>
<td>4.40</td>
</tr>
<tr>
<td>AVERAGE</td>
<td>T – TEST</td>
<td>-1.66</td>
</tr>
</tbody>
</table>
### TABLE 2A: AGEING EFFECT (DESCRIPTIVE)

#### SAMPLE 1: APPOINTMENT

<table>
<thead>
<tr>
<th>AGE GROUP</th>
<th># SUBJECTS</th>
<th>AVERAGE</th>
<th>ST.DEV.</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 21</td>
<td>2,360</td>
<td>21.17</td>
<td>17.58</td>
</tr>
<tr>
<td>21 – 30</td>
<td>4,570</td>
<td>20.17</td>
<td>17.79</td>
</tr>
<tr>
<td>31 – 40</td>
<td>4,772</td>
<td>18.42</td>
<td>17.03</td>
</tr>
<tr>
<td>41 – 50</td>
<td>3,079</td>
<td>19.63</td>
<td>17.11</td>
</tr>
<tr>
<td>51 – 60</td>
<td>2,941</td>
<td>19.17</td>
<td>17.11</td>
</tr>
<tr>
<td>61 – 70</td>
<td>5,049</td>
<td>19.32</td>
<td>16.77</td>
</tr>
<tr>
<td>71 – 80</td>
<td>3,738</td>
<td>18.83</td>
<td>15.90</td>
</tr>
<tr>
<td>≥ 81</td>
<td>704</td>
<td>19.84</td>
<td>17.58</td>
</tr>
</tbody>
</table>

#### SAMPLE 2: NUMBER

<table>
<thead>
<tr>
<th>AGE GROUP</th>
<th># SUBJECTS</th>
<th>AVERAGE</th>
<th>ST.DEV.</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 21</td>
<td>1,080</td>
<td>25.72</td>
<td>20.83</td>
</tr>
<tr>
<td>21 – 30</td>
<td>2,012</td>
<td>25.66</td>
<td>21.01</td>
</tr>
<tr>
<td>31 – 40</td>
<td>1,996</td>
<td>24.49</td>
<td>20.69</td>
</tr>
<tr>
<td>41 – 50</td>
<td>1,567</td>
<td>27.23</td>
<td>21.52</td>
</tr>
<tr>
<td>51 – 60</td>
<td>1,483</td>
<td>24.42</td>
<td>19.94</td>
</tr>
<tr>
<td>61 – 70</td>
<td>2,906</td>
<td>24.24</td>
<td>19.10</td>
</tr>
<tr>
<td>71 – 80</td>
<td>2,166</td>
<td>25.04</td>
<td>20.23</td>
</tr>
<tr>
<td>≥ 81</td>
<td>422</td>
<td>25.75</td>
<td>19.14</td>
</tr>
<tr>
<td></td>
<td>SAMPLE 1</td>
<td></td>
<td>SAMPLE 2</td>
</tr>
<tr>
<td>-------</td>
<td>----------</td>
<td>-------</td>
<td>----------</td>
</tr>
<tr>
<td></td>
<td>LEVENE</td>
<td>t-TEST</td>
<td>LEVENE</td>
</tr>
<tr>
<td>10-20 vs. 21-30</td>
<td>NR</td>
<td>R*</td>
<td>NR</td>
</tr>
<tr>
<td>21-30 vs. 31-40</td>
<td>R</td>
<td>R</td>
<td>NR</td>
</tr>
<tr>
<td>31-40 vs. 41-50</td>
<td>NR</td>
<td>R</td>
<td>R*</td>
</tr>
<tr>
<td>41-50 vs. 51-60</td>
<td>NR</td>
<td>NR</td>
<td>R</td>
</tr>
<tr>
<td>51-60 vs. 61-70</td>
<td>NR</td>
<td>NR</td>
<td>R</td>
</tr>
<tr>
<td>61-70 vs. 71-80</td>
<td>R</td>
<td>NR</td>
<td>R</td>
</tr>
<tr>
<td>71-80 vs. &gt;80</td>
<td>R</td>
<td>NR</td>
<td>NR</td>
</tr>
</tbody>
</table>

* $\alpha=0.05$ and ** $\alpha=0.10$
TABLE 3A: AGING EFFECT BY GENDER GROUP

<table>
<thead>
<tr>
<th>AGE GROUP</th>
<th>APPOINTMENT NUMBER</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;31 vs. 31-60</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MALE</td>
<td>F = 2.76 (0.09)</td>
<td>T = 4.26 (0.00)</td>
<td>F = 5.22 (0.02)</td>
<td>T = -1.56 (0.11)</td>
<td></td>
</tr>
<tr>
<td>FEMALE</td>
<td>F = 4.66 (0.03)</td>
<td>T = 3.84 (0.00)</td>
<td>F = 1.53 (0.21)</td>
<td>T = 2.32 (0.02)</td>
<td></td>
</tr>
<tr>
<td>&lt;31 vs. OLDER</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MALE</td>
<td>F = 7.59 (0.00)</td>
<td>T = 4.31 (0.00)</td>
<td>F = 0.30 (0.58)</td>
<td>T = -0.48 (0.63)</td>
<td></td>
</tr>
<tr>
<td>FEMALE</td>
<td>F = 31.91 (0.00)</td>
<td>T = 2.95 (0.03)</td>
<td>F = 20.69 (0.00)</td>
<td>T = 3.16 (0.00)</td>
<td></td>
</tr>
<tr>
<td>31-60 vs. OLDER</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MALE</td>
<td>F = 1.30 (0.25)</td>
<td>T = -0.08 (0.92)</td>
<td>F = 11.24 (0.00)</td>
<td>T = 1.34 (0.17)</td>
<td></td>
</tr>
<tr>
<td>FEMALE</td>
<td>F = 16.53 (0.00)</td>
<td>T = -0.88 (0.37)</td>
<td>F = 15.60 (0.00)</td>
<td>T = 1.01 (0.30)</td>
<td></td>
</tr>
</tbody>
</table>

(p-value between brackets).
### TABLE 3B: GENDER EFFECT BY AGE GROUP

<table>
<thead>
<tr>
<th>MALE VS. FEMALE</th>
<th>APPOINTMENT</th>
<th>NUMBER</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LEVENE</td>
<td>T-TEST</td>
</tr>
<tr>
<td>10 TO 30 YEARS OLD</td>
<td>F = 3.51 (0.06)</td>
<td>T = 1.95 (0.05)</td>
</tr>
<tr>
<td>31 TO 60 YEARS OLD</td>
<td>F = 5.19 (0.02)</td>
<td>T = 0.87 (0.38)</td>
</tr>
<tr>
<td>61 AND OLDER</td>
<td>F = 0.00 (0.95)</td>
<td>T = 0.18 (0.85)</td>
</tr>
</tbody>
</table>