Does Money Illusion Matter?  
A Theoretical and Computational Model

Niccolò Zaccaria

Working Paper N. 25/2019

The findings, interpretations, and conclusions expressed in the working paper series are those of the authors alone. They do not represent the view of Dipartimento di Scienze per l’Economia e l’Impresa
Does Money Illusion Matter?: A Theoretical and Computational Model

Niccolò Zaccaria†
September 3, 2019

Abstract

We based our work mainly on ‘Does Money Illusion Matter?’, by E. Fehr and J. R. Tyran (AER 2001) in which the authors show experimental evidence of the presence of money illusion within subject groups. We build up a model which provides a formal and mathematical framework of the experiment design, and it is in principle able to explain subjects’ behaviour within the experiment. Once we had analysed the dynamic properties of our model, we have run numerical simulations in order to see whether we were able to get the same pattern found by the authors. It turns out that not only our model is able to give a theoretical justification of the results found in the lab, but moreover it is able to replicate the experimental series found by the two couple of authors, up to a certain degree of fitness. Then we used a replication of the original experiment by L. Petersen and A. Winn, ‘Does Money Illusion Matter?: Comment’, (AER 2014), to test for robustness of our model.

Money illusion implies that agents’ behaviour depends on whether the same objective situations is represented in nominal or real terms. Mainstream economics assumes that «money is just a veil», which implies homogeneity of degree zero for demand and supply function. Whether people in fact are actually able to perceive the veil of money is just an empirical question. The authors indeed tried to look for empirical evidence of the ability, or inability, of people to see through the money veil. They approached this issue in an experimental environment, and found relevant and significant results, suggesting that money illusion matters.

Money illusion is an old fashioned topic, since it implies a certain lack of rationality which is somehow alien to most of the economists. Or at least, this was the dominant idea that came out from the rational expectation revolution. If we want to trace a brief description of money illusion history, we may start from a long ago, since it was I. Fisher one of the first economists to concern about money illusion, in 1928. He wrote a book simply called ‘The Money Illusion’, whose main purpose was to make the common audience aware of the possible harms coming from inflation, mostly in terms of a loss in their purchasing power. Money illusion was not still a research topic, but some sort of story tale about inflation that common people might safely ignore, but surely something that the representative agent had perfect knowledge of.

Then an important contribution came from F. Modigliani and R. A. Cohn (1979), who hypothesized that the stock market suffers money illusion, discounting real cash flows at nominal interest rates. Such hypothesis gave birth to an interesting research field in finance, aimed to investigate the actual role played by money illusion. Indeed, this article lead to several works aimed to empirically test the eventual role of money illusion, such as Vuolteenaho 2005 and Wang 2009.

Nonetheless, these contributions mainly remained within the finance field. Economic research did not take into account money illusion until the end of the nineties, when behavioral economics was spreading throughout the departments, and regained this topic.

†Università degli Studi di Firenze, DISEI, niccolo.zaccaria@stud.unifi.it. I am deeply thankful to Giorgio Ricchiuti, who provided insight and expertise that made this work come true, and to Leonardo Boncinelli and Nicola Doni, who immediately made them disposable to improve the quality of this research through comments and suggestions. Although, any errors are my own and should not tarnish their esteemed reputation.
As a matter of fact, one of the first and most relevant work on this topic was the article ‘Money Illusion’, by E. Shafir, P. Diamond and A. Tversky in 1997. Their aim was to show through survey questions that people often think about economic transactions in both nominal and real terms, and that money illusion arises from an interaction between these representations, which results in a bias toward a nominal evaluation. Furthermore they provided one of the first example of a twist on an already existing model, that is the Solow efficiency wages model (1979), implementing money illusion through a parameter.

Aside from the very results of this work, its actual merit was to give birth to a rich stream of literature about money illusion. In particular, experimental economics revealed itself to be extremely useful when it comes to test whether people can perceive the veil of money.

Although the numerous experimental works about money illusion, a microfoundation for this phenomenon is still missing. Actually, there are few models that try take into account money illusion. Some attempts have been made to embed money illusion in already existing model, as it is the case of Shafir, Diamond and Tversky. But there is still no general theoretical framework for money illusion. The solely attempts to model money illusion tout court come from a more financial approach, like the one made by S. Basak and Y. Hongjun (2010) in ‘Equilibrium Asset Prices and Investor Behaviour in the Presence of Money Illusion’.

Money illusion seems to be a psychological phenomenon difficult to be universally defined. Rather it seems easier to define it case by case, according to the economic environment we decide to place ourselves in. As a matter of fact, we have tried to follow this very path, and to define money illusion in a specific case coming from the experimental literature. This work shall not be seen as a general microfoundation of money illusion; rather it may be considered as a behavioral microfoundation of money illusion up to some specific extent.

1 Introduction

This work is based mainly upon the article Does Money Illusion Matter?, by Ernst Fehr and Jean-Robert Tyran, published in the American Economic Review in 2001, henceforth FT. In this paper the authors show that a small amount of individual level money illusion may cause considerable aggregate nominal inertia after a negative nominal shock. Moreover, they found evidence for an asymmetric effect of a positive money shock. Still, they assume this effect to be a consequence of money illusion.

The authors’ intuition is that money illusion has two components:

1. Direct effects: people making errors in optimization.
2. Indirect effects: people expecting others to make biased decisions and react strategically to this.

In an interactive situation the failure of some agents to fully adjust to the nominal shock will, in general, provide incentives for other agents to not fully adjust to shock either. There must be a snowball effect that causes less than full adjustment for an extended period of time. This result may be due to the money illusion rendering price expectations very sticky after a negative shock. And, under conditions of strategic complementarity, price stickiness seems to stronger affect players’ behaviour.

Hence, authors believe that money illusion may offer an alternative reason why prices can show evidence of sluggishness.

Subsequently the publication of FT, two other scholars, L. Petersen and A. Winn, here after PW, strongly criticized FT’s findings. Their reply to FT took form as an article published in the American Economic Review in 2014, Does Money Illusion Matter?:Comment. They disagreed about FT’s claim of describing the experimental evidence in terms of money illusion. According to PW the nominal inertia is not caused by money illusion, which has only slight second order effects. They believe that inertia is caused by the cognitive challenge due to subjects’ decision task. The duration of nominal inertia depends primarily on participants’ best response function, not the prevalence of money illusion. Hence, their critique was mainly about the interpretation of the results; not about the results themselves. Indeed, they replicated the experiment, keeping the design unchanged, and found similar results to FT.

On one hand, it is not in our interest to enter this debate. Fehr and Tyran also replied (Fehr and J. R. Tyran 2014), carrying on their own idea about the role of money illusion. On the other
hand, we acknowledged that none of the two couple of authors have tried to address this topic in a formal manner. Indeed, Fehr and Tyran provided a rich description of the results, explaining in detail which role the money illusion is supposed to play. However, a precise model that could be able to explain ex ante the findings behind money illusion is still missing.

Hence we developed a plausible model that could be able to explain direct and indirect effects of money illusion. Such model is built directly on the grounds of the experimental design provided by FT, that is, we assume $n$ players to play a pricing game in a monopolistic competition environment. Our model entails that each player, at each time $t$, decides which price he wants to set following a forecasting rule. Such forecasting rule is composed by two terms: a belief about which will be the current average price, and the last period average price. We assume money illusion to be given by the weight that each players give to the former or the latter term.

On one hand the belief captures the absence of money illusion, that is, it as long as players give more weight to such term, we are assuming them to be more precise in forming their belief, and therefore we are assuming less “illusion”. On the other hand the latter captures the stickiness of price dynamics, that is players keep following the previous price.

First, we replicated FT findings. Then, we used it to replicate PW results. Actually we used PW results as a sort of robustness test for our model. Our objective was to see whether the model was able to provide a good degree of fitness for both the experimental works.

Obviously our model entails money illusion, at least in its construction. And we used it to replicate PW data, which did not suppose the presence of money illusion, at least in their initial intention. However our aim is to build a model able to explain money illusion. Therefore, on the top of this, we are implicitly relying on FT interpretation of the results, that is money illusion may actually be present in a certain environment, and may cause considerable nominal inertia.

2 Money Illusion at Individual and Aggregate Level

Firstly we give a rhetorical description about how money illusion could affect players’ behaviour, eventually ending up to spread throughout the whole system.

Agents were not traditionally believed capable of being fooled just by a change in nominal magnitude. What should actually matter is the real incentive structure of an economy, that is, as long as the objective situation remains unchanged, agents’ decision is supposed not to change either. So, objective function is not supposed to change depending on nominal magnitude, as long as agents can see through the veil of money. Either, agents should perceive that purely changes in nominal term do not affect their opportunity set.

FT work shows evidence that these two assumptions seem to be violated in lab environment. Money illusion at the individual level suggests that the preferences of many people, as well as their perceptions of the constraints, are affected by nominal values. Though, even more interesting is the money illusion at the aggregate level. Assume the population of agents to be completely rational, and immune to money illusion, but for one member. We may show with a simple example and a trivial use of backward induction, that money illusion may cause a paradoxical result.

First let a $n$-players dynamic game be given, with $n = 2$. So we have player 1 and player 2. The game is described as follows:

Both agents may choose between two actions, a rational one, $R$, and a bounded-rational action $U$. Assuming rationality, there is no reason why the game should not end in $(R, R)$. But now say that player 2 is a money illusion prone individual, and he, or she, is unable to screen the rational action from the irrational one. And say that the action $U$ may even look more attractive than $R$ to him, since it leads to higher nominal payoffs. So assume that player 2 will always play $U$. Then, in regular games players have no reason to doubt about opponents’ rationality, but let’s forget for a moment regular games. Say that player 1 may have a suspicion about player 2’s bounded rationality. Let’s say that there is a probability $p$ that player 2 is bounded rational, according to 1. Then 1 may still move rationally and play $R$, or he could go for the prone-illusion action too, in order to protect himself. Then the game would end in $(U, U)$. This is an example of how bounded rational individuals may drive also rational ones to act in an irrational way.

See that player 1 may still behave in a rational way, as long as the expected value of the loss given by the node $(R, U)$, weighed by probability $p$, is greater than other possible wins. This would also be coherent, from a behavioural point of view, with loss aversion theory. As the feared loss
caused by bounded rational actions from bounded rational individuals becomes larger, also rational agents would be crowded out and nudged to play safely, even following illusion prone individuals.

Now take the same game with \( n \) players, and say that only the \( n \)-th player is myopic, and playing constantly \( U \). Then a cascade effect will occur. The \( n - 1 \)-th player will also play \( U \), if he fears player \( n \) to be illusion prone. And so the \( n - 2 \)-th player will play \( U \) as well, as long as he fears other players to be prone to money illusion. And through backward induction the money illusion would affect every player in the game, leading to an inefficient outcome for everyone. Still, \( n - 1 \) players may have played according to rational assumption.

This obviously would depend on the payoffs involved in the game. Although, we could reasonably think of a wide range of payoffs leading to such cascade effect. The only feature we need is a large gap between possible wins and losses such that rational agents will badly fear an irrational action from anyone else in the game. Hence, they could decide to act prudently, and protect themselves from possible losses due to illusion prone individuals.

This is not a rigorous proof of how money illusion could spread through a sequential game. We will not even prove anything about possible equilibria out of this game. This is meant to be just a hint about how money illusion could act at the aggregate level. We believe that the cornerstone of this topic is not about how many agents are affected by money illusion, rather than it is about how many agents believe that others may be affected by money illusion.

We now present the experimental design that has been followed by the authors. Then we provide a plausible model that could explain the findings of the experiments.

### 2.1 Experimental Design

The experiment lasts \( 2T \) periods, and consists in a \( n \)-players pricing game with strategic complementarity and a unique equilibrium. The players have to decide simultaneously their nominal prices in each period, and were free to change their prices at no cost. Then, at the end of each period, they are informed about the payoff they have gained, and previous earnings. The payoff function used throughout the whole experiment remains constant:

\[
\pi_i = \pi_i(P_i, \bar{P}_{-i}, M),
\]

where \( P_i \) is the price that players have to choose, \( \bar{P}_{-i} \) is the average price of the \( n - 1 \) players group, and \( M \) is the money supply. Everything is evaluated at time \( t \).

In order to choose the price and maximize future payoff they are given a table containing all the possible payoffs as a function of their price and of the average price of the \( n - 1 \) group. This feature makes playing a best reply for a given expectation about the other players’ average price particularly easy. In the first \( T \) periods of the experiment money supply is given by \( M_0 \). Then a fully anticipated monetary shock is implemented by reducing the money supply to \( M_1 \). The shock

---

\[^1\text{You may find the complete functional form of the payoff function used by the authors in the Appendix.}\]
Table 1: The treatments used throughout the experiments, combining the representation of the payoffs and the nature of players’ opponents.

<table>
<thead>
<tr>
<th>Payoffs</th>
<th>Opponents:</th>
<th>Human</th>
<th>Computerized</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal</td>
<td>NH</td>
<td>NC</td>
<td></td>
</tr>
<tr>
<td>Real</td>
<td>RH</td>
<td>RC</td>
<td></td>
</tr>
</tbody>
</table>

is represented by a change in the values contained in the payoff matrix that the players are given. This makes a big difference according the treatment the players are assigned to. In total we have four treatment conditions, which differ along two dimensions.

The first dimension concerns the framing of the situation, whether the payoffs were represented in real or nominal terms. Since the nominal payoff is given by \( P_{-i} \cdot \pi_i \), they had to divide the values contained in their payoff matrix by \( P_{-i} \).

The second dimension concerns the fact whether our experimental subjects face \( n-1 \) preprogrammed computerized players or whether they face \( n-1 \) other human subjects. The response rule of the computers is given by the best replies of the computers based on (1). Subjects know that they play against computers, which are programmed to play best reply to humans, therefore there is no strategic uncertainty, and no need for for form expectations for others’ behaviour. Moreover, since the computers play best replies, their behaviour rules out any money illusion. On the other hand, in the condition with human opponents each subject faces the task of forming expectations about the other players’ price choices. This necessarily involves a guess about the extent to which other players are affected by money illusion.

Hence combining the two classes of treatments we have four different cases:

The payoff function is characterized by the following properties.

**Proposition 1**

(i) The payoff function is homogeneous of degree zero in all of its arguments \( P_i, P_{-i}, \) and \( M \).

(ii) The best reply function is weakly increasing in \( P_{-i} \).

(iii) The equilibrium is unique for any \( M \).

(iv) The equilibrium is the only Pareto efficient point in the payoff space.

(v) The equilibrium can be found by iterated elimination of weakly dominated strategies.

Proposition [i] and [iii] implies the existence of a unique money-neutral equilibrium \( P^*_i \) for \( i = 1, \ldots, n \). They were implemented in order to analyze the impact of money illusion on the adjustment of an economy without the possible crowding out effect of multiple equilibria. Although the proof of these propositions is beyond the scope of this paper, we can show that they imply the neutrality of money. See that a money change in \( M \) from \( M_0 \) to \( \lambda M_0 \), leaves real payoff unaffected as long as prices change to \( \lambda P_i \) and \( \lambda P_{-i} \). Moreover if \( P'_i \) is a best reply to \( P_{-i} \) at \( M_0 \), then also \( \lambda P'_i \) is a best reply to \( \lambda P_{-i} \) at \( \lambda M_0 \). Everything is just rescaled, but the relation between the variables remain unaffected. Hence \( \lambda P^*_i \) is the post-shock equilibrium.

Proposition [ii] introduces the strategic complementarity in the game, and Proposition [iv] implies that any deviation from the equilibrium involves a real income loss for the players.

Finally Proposition [v] means that there is a method for finding the equilibrium that works exactly in the same way in the real, as well as in the nominal frame. In the real treatments a weakly dominated strategy \( P_i \) has (weakly) smaller payoffs value at any level of \( P_{-i} \). On the other hand in the nominal treatments only the magnitudes of strategies change, but not their order condition. Therefore to eliminate weakly dominated strategies in either treatments, subjects only need to eliminate those strategies that have weakly smaller (real or nominal) payoff numbers at any given level of \( P_{-i} \).

---

2The proof of this proposition must be looked for in the Appendix of ft2001.
2.2 A Plausible Model

In this section we are presenting a behavioral microfoundation of money illusion. Throughout this model, we will be able to replicate the findings reported by FT and PW.

We assume that for each player there is a forecasting rule to choose the price in each period, which is given by the following convex combination, given \( \omega \in (0, 1) \):

\[
P_{i,t} = \omega X_{i,t} + (1 - \omega) P_{t-1},
\]

where \( P_{t-1} \) is the last period average price. Therefore players take into account the last period average price, and a random variable \( X_{i,t} \) in order to choose their price. \( X \) represents the subjective distribution about what will be the other players choice in the current period \( t \). There is an average value of \( X \) for each \( t \), which is the actual average of players’ prices that will occur at the end of the period. That is

\[
E_t[X_t] = P_t,
\]

The fact is that players cannot know the realized value of \( X_t \) until the period ends. Hence they cannot rely on the unconditional expectation: they must rely on a conditional expectation of \( X_t \).

Therefore they will forecast \( X_t \) conditional to the information set \( I \) available at time \( t \). The information set at time \( t \) contains only the values of \( P \) and \( M \) up to time \( t-1 \). Therefore each player needs to form an expectation about average price at date \( t \), and this expectation is given by the distribution of \( X \). Formally the information set is made as

\[
I_t = \{ M_t, M_{t-1}, P_{t-1} \}.
\]

It is as if they were to guess which price will be the most chosen by the others, and with which probability. We suppose players to know the value of \( M \) only for periods \( t \) and \( t-1 \), that is we are assuming that players’ memory cannot go further than one lag. We are analysing the effects of a money shock; we are not interested in testing the players’ behavior as a continuous function of money, at least not here.

The random variables \( X_i \) are i.i.d. for each players; we assume that on average all players will have the same degree of precision among their subjective distribution. We assume the following normal distribution:

\[
X_{i,t} \sim N(\mu + \theta_t, m_t).
\]

The distribution has mean value \( \mu + \theta_t \), where \( \mu \) would be the mean of an unbiased distribution where money illusion takes no role. \( \mu \) is nothing more than the exogenous equilibrium price of the experiment. So the conditional expectation of the random variable is

\[
E_t[X_{i,t}|I_t] = \mu + \theta_t,
\]

such that the variance and the term \( \theta \) will capture the direct effects of money illusion.

We are assuming that players have at least a glance of how the equilibrium is made, but due to a bias \( \theta \) they may not be able to reach it, or even get close to it. We define \( M_t \) as the current quantity of money present in our economy. Then the parameter \( m_t \) is defined as the ratio between money supply in the current period and in the previous one:

\[
m_t = \frac{M_t}{M_{t-1}}.
\]

We can see this parameter as the precision of players’ distribution. As long as quantity of money remains constant through time, the precision (and the variance) of the distribution equals one. If there is a negative shock instead, the ratio becomes larger, and the precision drops off, giving rise to price sluggishness. On the other hand, if there is a positive shock, the ratio decreases, and the precision gets larger. This is supposed to boost price adjustment, which is coherent with the results of the experiment about the asymmetric effects of negative and positive money shocks.

Finally, \( m_t \) enters also in the bias of the distribution, i.e. \( \theta_t \). This error is defined as the natural logarithm of \( m_t \), such that when money remains constant, this term simply goes away,

\[
\theta_t = \beta \log \left( \frac{M_t}{M_{t-1}} \right),
\]
with $\beta$ being a switch parameter which plays a crucial role whether the shock is positive or negative. Indeed, on the grounds of the results by FT, the shock is understood to have asymmetric effects whether it is negative or positive: on one hand a negative shock causes a considerable amount of nominal inertia, on the other hand the positive shock seems to have only slighter effects on the adjustment process. The reason seems to be that subjects feel no hesitation whenever they have to raise their nominal price due to a positive shock. Rather, money illusion seems to imply that people feels a lot more of hesitation when they have to lower their price. Therefore, on the top of this we define $\beta$ as a switching parameter

$$\beta = \begin{cases} 
1 & \text{if } M_t \geq M_{t-1} \\
0 & \text{if } M_t < M_{t-1}
\end{cases} .$$

(9)

Such definition allows us to get rid of the bias $\theta_t$ in the case of a positive shock. This may seem just an arbitrary definition coming out of the blue. Nonetheless such definition of $\beta$ is perfectly coherent with the hypothesis of an asymmetric effect of a money shock, and moreover, as you may find in the Appendix, the presence of $\beta$ finds no legitimacy in the case of a positive shock. In Section 4 we are going to show properly why we took such decision.

The idea behind all these handling of the money supply is that players rely heavily on the nominal framing of the game. Since they have no clue about others’ reasoning, the only thing upon which they can build their expectations is the quantity of money in the economy, which happens also to be the only visible and certain value at time $t$, apart from the previous average price. Therefore they expect also other players to think likewise they are doing, and since the only information which is common knowledge is money supply, everybody ends up to base their subjective distribution on money supply. The fact that in the distribution is present the true equilibrium price represents the idea that although they could not fully reach the equilibrium, players are aware of its presence, and even of its shape as long as money change does not interfere with their guessing of the price.

Having a random variable with a certain bias and a certain noise has little to do with strategic complementarity. The idea is that changes of money enter direct in players computation, causing them a probability to make some mistakes. Indeed, what makes the price stickier is the parameter $\omega$, that is, how much weight players attach to the previous value of the average price. It gives the level of how confident players are about stickiness of others’ expectations. A lower $\omega$ tells us that players are pretty confident that the others will not stick to the previous price. A higher $\omega$ implies that the subject prefers to draw the price from his subjective distribution, which mostly rely on money supply. Later on we will define the parameter $\omega$ as a confidence parameter. Moreover, we may believe it to be negatively related to money illusion: as long as we assume money illusion to cause some nominal inertia, a higher omega causes a quicker adjustment, and less sluggishness.

Therefore, the equation that we are using to make up single choices about the price is

$$P_{i,t} = \omega X_{i,t} + (1 - \omega) \sum_{i=1}^{n} \frac{P_{i,t-1}}{n},$$

(10)

such that at each time $t$ the average price is computed as it follows

$$\bar{P}_t = \sum_{i=1}^{n} \frac{P_{i,t}}{n}.$$  

(11)

Now we move to analyse the dynamical properties of the dynamic equations we have described so far. First we look for the steady states of our system.

Let’s consider a fixed point for individual $i$. We name $f$ the function of (2), and we give it a proper definition on its domain and range:

$$f: \Omega \rightarrow \mathbb{R}, \quad f: X_{i,t} \mapsto P_{i,t},$$

such that $\Omega$ is the stochastic space the subjective distribution $X_i$ belongs to. Then any fixed point for player $i$ must be a price such that, given $P_{i,t}$, the following must hold:

$$P_i = f(P_i).$$
See that as long as money remains constant through time the distribution \( X_i \) is not really time dependent. Then let’s explicit the very definition of \( f \):

\[
P_{i,t} = \omega X_{i,t} + \frac{1 - \omega}{n} \sum_{i=1}^{n} P_{i,t-1}
\]

Take the price for individual \( i \) off the average, and get

\[
P_{i,t} = \omega X_{i,t} + \frac{1 - \omega}{n} \sum_{j \neq i} P_{j,t-1} + \frac{1 - \omega}{n} P_{i,t-1}.
\]

Then any fixed point for individual \( i \) will be a price such that \( P_{i,t} = P_{i,t-1} = P_i \). Therefore our dynamic equation evaluated at such fixed point becomes

\[
P_i = \omega X_{i,t} + \frac{1 - \omega}{n} \sum_{j \neq i} P_{j,t-1} + \frac{1 - \omega}{n} P_i
\]

(12)

For sake of brevity call \( \frac{1 - \omega}{n} = \alpha \), then

\[
P_i(1 - \alpha) = \omega X_{i,t} + \alpha \sum_{j \neq i} P_{j,t-1},
\]

such that we can write the equation for the steady state for individual \( i \) as

\[
P_i = \frac{\omega X_{i,t} + \alpha \sum_{j \neq i} P_{j,t-1}}{1 - \alpha}.
\]

(13)

**Proposition 2** If the subjective distributions \( X_i \) are identically and independent distributed among the players, then a steady state of (2) exists for each individual \( i \). Furthermore, if such steady state exists, it is given by the following strategy

\[
P^* = X_{i,t},
\]

(14)

for any \( i \in I \).

**Proof.** Take the previous definition of the steady state for only individual \( i \). As long as \( X_i \) are i.i.d., we may safely state that the way in which the expectation of other players’ price is chosen must be identical among all players. Then it follows that the equilibrium strategy given by the eventual steady state must not differ among different players. We may rewrite it as \( P_i \), or simply \( P^* \), since it must hold equal for everyone. Therefore we have

\[
P^* = \frac{\omega X_{i,t} + \alpha(n - 1)P^*}{1 - \alpha}
\]

for any \( i \). Then

\[
P^*(1 - \alpha) + \alpha(1 - n)P^* = \omega X_{i,t}
\]

\[
P^*(1 - n\alpha) = \omega X_{i,t}
\]

\[
P^* = \frac{\omega X_{i,t}}{1 - n\alpha} = \frac{\omega X_{i,t}}{1 - n\frac{\omega}{n}} = X_{i,t},
\]

that is, each player will set its price equal to its subjective distribution \( X_{i,t} \).

What is the meaning of the above proposition? Although it seems rather abstract, it gives a formal proof of a simple and reasonable idea. As long as agents are rational, and therefore no signs of money illusion appear, each of them is perfectly able to reach the equilibrium of the game. How come? Simply playing their subjective idea that everybody is going for the equilibrium, which is nothing more than the expected value of \( X \). Hence, apart from the intrinsic randomness of a
stochastic distribution, as long as money plays no role, everyone is on average able to guess which price will be chosen by the majority. And, eventually the most salient fact, this result shows us that if money is constant then everybody is aware that everybody else is aware of the pointless role of money. It means that the fact that money cannot perturb the game from its equilibrium is common knowledge among the players.

From this proposition an other interesting result may follow.

**Corollary 3** The expected value of $X$ is also a fixed point, i.e.

$$E[X_t] = P_t = P^*.$$  \hspace{1cm} (15)

**Proof.** Take the definition for the individual price $P_{i,t}$ given in (2). Then the current average at time $t$ is given by

$$P_t = \frac{1}{n} \sum_{i=1}^{n} P_{i,t} = \frac{1}{n} \sum_{i=1}^{n} [\omega X_{i,t} + (1 - \omega)P_{t-1}]$$

$$P_t = \frac{\omega}{n} \sum_{i} X_{i,t} + (1 - \omega)P_{t-1}$$

$$P_t + (\omega - 1)P_{t-1} = \frac{\omega}{n} \sum_{i} X_{i,t}$$

Now suppose that there is a fixed point such that $P_t = P_{t-1} = P^*$. Then we would get

$$P^* = \frac{1}{n} \sum_{i=1}^{n} X_{i,t}$$  \hspace{1cm} (16)

which is the average of the subjective distributions. But $X_{i,t}$ is drawn from a normal, therefore as the sample size grows up to infinity, by the Central Limit Theorem such average is equal to the expected value of the random variable. That is as $n \to \infty$

$$\frac{1}{n} \sum_{i=1}^{n} X_{i,t} \to E[X_t],$$  \hspace{1cm} (17)

$P^*$ is a fixed point. \hspace{0.5cm} \blacksquare

Therefore $P^*$ is also a fixed point, and, as sample size grows, it is also an unbiased estimate of the expected value of $X$. But there is actually no need to trouble the Central Limit Theorem. See that if we evaluate the system at the steady state found in the previous proposition, and you take the average of the individual price at any date $t$, you get

$$P = \frac{1}{n} \sum_{i=1}^{n} P_{i,t} = P^*.$$  \hspace{1cm} (18)

Hence equation (16) and (18) are describing the same fixed point. From the previous proposition, if we evaluate the system at $P^*$, it follows that the average of such fixed points, is simply $P^*$, since it is equal throughout all players. Eventually, this corollary tells us nothing new. Which are the actual consequences of these claims on the individual choice of $P_{i,t}$?

Suppose the existence of this equilibrium is common knowledge, which simply requires $X_i$ i.i.d. for each $i$. This would entail everybody to play according to their distributions, having no suspicion that someone could deviate, or not find, the equilibrium. Then everybody would play $E[X_{i,t} | I_t] = \mu + \theta_t$. However as long as money remains constant $\theta_t$ disappears, and everybody would end up playing $P_{i,t} = \mu$, which turns easily out to be the equilibrium described at the beginning of this section. Hence, we have proved so far that money is not able to prevent players to reach the equilibrium. Unless it does change.

What if money changes from $t$ to $t+1$? Nothing weird, there is an only consequence, though relevant. People cannot reach the equilibrium now, due to the presence of the mistake $\theta_t$ within their distributions. Hence as long as $\theta_{t+1} > 0$, then $E[X_{i,t} | I_t] \neq E[X_{i,t+1}]$. Nonetheless the equilibrium of the game was supposed to be exogenously given, therefore any deviation at $t+1$ from $\mu_a$ will incur in a deviation from the equilibrium.
3 Simulations to Replicate the Experiments

In this section we use the model presented so far to replicate the findings of the authors, both in negative and positive shock cases.

The authors, in order to test the robustness of their experiment, tested their hypothesis about money illusion using first a negative shock, then a positive one. We replicate their data in both cases.

We have stuck to the framing of the experiment, with $T = 40$ and $n = 4$ players. Also payoff functions remains the same, the only thing that we have changed, for sake of simpleness, is that now $P_i$ goes into the computation of the $n$ players average $\bar{P}$. Since no actual human players were involved in this computation, we believe that this twist would make no harm.

Since we are mostly interested in the indirect effects of money illusion, we will rely only on the Nominal Human opponents treatment. Doing so we can also analyse the direct effects of money illusion, that is errors in optimization.

In each period players chose their subjective price following a forecasting rule, then at the end of the period prices of all players are summed and the resulting average is considered as the actual price of that period.

3.1 Negative Shock

First we show the experimental evidence found from FT in 2001. As we can see from Figure 2, there is little presence of noise around the pre-shock equilibrium, though nothing strong enough to make the game deviate from the steady state. Then at time $T = 20$ the money shock was implemented, and we can see that the average price leaps downward, showing strong evidence of stickiness. According to FT, this stickiness is due to the presence of money illusion. Beware that we are not assuming subjects to be prone to money illusion, and unable to perceive the difference after the shock. Though, we believe that this could be only one of two reasons: mainly we believe that subjects react so slowly and sticky since they were expecting others to be prone to money illusion. If money illusion totally had played no role, we should have seen the average price immediately adjusting in the case of a positive shock to the new equilibrium.
Figure 3: Average price dynamics with experiment level of money supply, and four different values of \( \omega \).

We have tried to replicate the experimental data letting the parameter \( \omega \) take several values, and we display below in Figure 3 the outcomes for \( \omega \) equal to 0.1, 0.3, 0.5 and 0.7. We implemented the same values of money supply \( M_0 = 42 \) and \( M_1 = 14 \) after the shock.

As you can see, we may already a similar pattern to the one plotted in Figure 2 for omega being between 0.1 and 0.3. Thus, we are going to replicate it again assuming \( \omega = 0.2 \), and to compare the replication with the experimental data. In Figure 4 we can see the authors’ results, juxtaposed to our replication. The thicker blue line represents the replicated series, and the red line represents the data collected by the authors. We can observe a certain degree of fitness between the two series.

We can see immediately that there are two features that the two series share: a certain degree of stickiness, and a certain degree of noise. Let’s see why. We can see from (2) that the second member of the equation on the right hand side, i.e. \((1-\omega)P_{t-1}\), is the one that drives the stickiness of price formation. The lower is \( \omega \), that is the higher is \( 1-\omega \), and the stickier is the average price. This implies that agents, disregarding their own subjective distribution, are relying mostly on the previous value of the average price. This tells us that there is a low level of confidence about other players’ precision in the game.

On the contrary, the noise of the series seems to be affected only by the money shift. Especially when \( \omega \) is higher, which means that players are relying mostly on their subjective distribution, and caring less about previous price value, we can see that the noise is so strong that prevents the players to reach the equilibrium. This happens since every distribution \( X_t \) strictly depends on the quantity of money: as the change from \( m_{t-1} \) to \( m_t \) increases, the bias of the distribution becomes larger, since the log is a strictly increasing function. So, even ignoring the precision of the distribution, we can already know that whatever form it may take, it will be biased by \( \theta_t \). The larger the difference between the money of the two periods is, the larger the bias will be. And, mostly, the
Figure 4: FT’s data compared to our model, with \( \omega = 0.2 \).

noise depends on the precision of the distribution, which is a decreasing function of money change from \( t - 1 \) to \( t \). Indeed, the larger the shock will be, the lower the precision, and the higher the variance, leading to a persistent noise. If we even supposed that the mean of the distribution were correct, the equilibrium would hardly be reached, since the series would just oscillate around it.

3.2 Postive Shock

Now we replicate again the experimental findings in the case of a positive money shock.

The authors changed only three things in the experimental design of the positive shock treatment:

1. They dropped the computerized opponents treatment.
2. They made the experiment last for \( T = 30 \).
3. Now the equilibrium price is 12.5 before, and 25 after the shock.

Anything else remained unchanged.

Likewise we kept the setting of the replication unchanged, except for the changes which were required by the new experiment.

In Figure 5 we display the results found from the authors FT.

If we stick to the value of \( \omega \) that we have used so far, i.e. \( \omega = 0.2 \), we get a bad description of the experimental data. Or, still we find the same trend, but the fit is poorly represented. You can see this in the Figure 15 of the Appendix. Instead we may make a further assumption here, which will find empirical validity in the next section. But firstly we would like to give a rhetoric description of what may be happening here.

We may think that, since we are in the presence of a positive shock, it seems unlikely that players rely mostly on the previous price, that is a low \( \omega \), as we did before. We believe it seems more plausible in this case to have a higher value of \( \omega \), that is, subjects should play according to their distribution, whose precision has now increased. The idea is that everybody should expect everybody else to have fewer hesitations in rising the price, and therefore following the positive shock. This is probably the reason why we got a better fit increasing players’ confidence in their own subjective distribution, that is we have tried with \( \omega = 0.6 \).
average price dynamics
authors results
pre-shock eq.
post-shock eq.
Experimental data with positive shock

Figure 5: FT’s experimental data with a positive shock.

We may even think of a possible quantitative relation involving of \( \omega \), and the sign of the shock. From what we have observed so far, we should expect a lower \( \omega \) with a negative shock, and a higher \( \omega \) with a positive one.

Indeed, if we plot together the authors’ results, and our replication with a different \( \omega \), that is \( \omega = 0.6 \), we have almost a perfect fit, as we can see in Figure 6.

4 The Role of Confidence as an Exogenous Parameter

An interesting feature would be to see all the possible outcomes of the average price dynamics, as a function of \( \omega \). Hence, we plotted a three dimensional graph in Figure 7 having on the three axis, respectively, the average price, time \( T \), and the values of \( \omega \). We kept the experimental parameters the same as the original ones from the authors.

Then we may wonder which is the \( \omega \) that better fits the experimental values, given these parameters. In order to answer to such question, we have done the following. We have taken \( \omega \) increasing by steps of 0.01, that is we took one hundred values of \( \omega \). Therefore we have the following sequence for \( \omega \)

\[
\{ \omega \}_k^K,
\]

ranging from 0 to 1, where \( k \) is an index going from 0 to \( K = 100 \).

We denote with \( \hat{P}_t \) the values found by the authors, that is the red line in the previous graph.

Then we compute the modulus of the distance between our replication and the experimental value for the average price. We have done this for each period \( t \), and for each of the \( K \) values taken by \( \omega \):

\[
x_{k,t} = |\hat{P}_t - T_t|.
\]

Actually we computed the following matrix, of dimension \( T \times K \):

\[
x_{T \times K} = \begin{bmatrix}
x_{1,1} & \cdots & x_{1,K} \\
\vdots & \ddots & \vdots \\
x_{T,1} & \cdots & x_{T,K}
\end{bmatrix},
\]

(21)
Figure 6: FT’s data about positive shock compared to our model, with $\omega = 0.2$.

Figure 7: Average price dynamics with negative shock, for $\omega$ going from 0 to 1, by steps of 0.01.
where each row represents the point-wise distance in each period between the experimental average price and our replication of the average price, for a single value of $\omega$. We computed the row average to get a single value for each time sequence, given a certain $\omega$:

$$\bar{x}_k = \frac{1}{T} \sum_{t=1}^{T} x_{t,k}$$

(23)

Finally, we can display in Figure 8 the dynamics of $\bar{x}_k$ as a function of the values taken by $\omega$. On the horizontal axis we have the values of $\omega$. On the vertical axis we have the average distance, computed with the $k$-th values of $\omega$.

Even if we may look for the point-wise minimum taken by the average distance, we are rather interested in finding a local neighbourhood in which the average distance is minimized. As we can see from the plot, this value looks to be around $\omega = 0.18$, for which the average distance is approximately 0.6. How should we interpret this results? Is it good or is it not? We may first consider it in absolute value: both series are fluctuating around the two equilibria, pre and post shock, which were given by a price of 18 and 6, respectively. Therefore an average distance of 0.6 in a series ranging roughly between 18 and 6 may be considered as an acceptable deviation from the original series. In Section A.4 of the Appendix you may see also the standard deviation between the two series. However we believe that average distance given by the modulus gives a more intuitive idea of the actual distance between the two series. We are talking about the dynamics of some price, hence it may be thought as reasonably close result having the average distance roughly equal to 0.6.

This is instead the case of a positive shock, still with $\omega$ going from 0 to 1. Now $T$ lasts only thirty periods, and money supply order is reverted. In Figure 9 is the 3d dynamics likewise we did for the negative case.

Figure 8: Average distance between FT’s data and our replication with negative shock.
We can see that, *ceteris paribus*, the positive shock shows a significant smaller amount of noise for the same values of $\omega$. This is perfectly coherent with the results obtained by the authors, as well as it is with the results driven from our model. All the paths of the average price in this case, as it is clearly evident from the plot, are much smoother than the ones from the negative shock case. This strengthens the hypothesis of an asymmetric effect between money shocks.

Furthermore if we plot the average distance as we did before, we get the result which is shown in Figure 10.

As we can see, there is not a clear minimum peak, already at first glance. Even though, for $\omega$ in the interval between 0.4 and 0.5, the average distance drops below one, and then it begins to grow up again. Therefore if we take $\omega$ belonging to this interval, we might get slightly better results.

5 Analysis of Robustness

We have relied our replication upon experimental data, therefore it would have been hard to decide how to modify them to test for robustness. We could have randomly changed the data, and see whether the same values of $\omega$ still minimizes the average distance. Or at least, whether the same neighbourhood of $\omega$ still could do the same.

Luckily, someone else came in our help, even with no intention. The work by Fehr and Tyran hit, positively and negatively, the community, because they find one of the first experimental evidence of the (likely) existence of money illusion. Obviously they received several critiques about their findings. One in particular came from L. Petersen and A. Winn (2014), as we already mentioned at the beginning. As a matte of fact, PW replicated the very same experiment of FT in order to disprove their results, or at least, their interpretations of the results. Indeed, they replicated the experiment, keeping the design unchanged, and they found data that, surprisingly or not, keep the same pattern of FT work.

This is the plot that we get for the replication of the negative shock. The only difference in PW design was the time length: they used thirty period rather than forty. Anything else remains unchanged. We get a very similar result as we got before. The distance is minimized in a
neighbourhood of $\omega = 0.2$. In this case the minimum peak is slightly shifted forward, keeping lower distances for greater values of $\omega$. Nonetheless, we can still claim with enough conviction that the value $\omega = 0.2$ gives also in this case a good fit of the data, as Figure 11 shows.

The result in the case of the positive shock is a bit different. Actually, with PW data it is even more precise. As you can see in Figure 12, the average distance is kept below one from values of $\omega$ ranging roughly from 0.25 up to 0.90. Instead, our replication of FT gave a distance lower than unity only for $\omega$ from 0.40 to 0.50.

6 Conclusion

Money illusion took several faces during the economic literature. On one hand only macroeconomics carried the weight of arguing about money illusion. On the other hand recent works, like the ones which this paper has took inspiration from, showed that money illusion can play a relevant role also in a strategic environment.

FT experiment proved money illusion to have some relevant consequences, although the authors did not come up with an a priori model. On the contrary, our model either gives a microfoundation for money illusion and is able to replicate with a good degree of fitness the experimental results. Of course we are aware of the several shortcomings our model suffers. First of all, the distinction between negative and positive shock has been implemented ad hoc, although in the previous section we showed the irrelevance of $\beta$ in the case of a positive shock. One may criticize that our model is not able to take independently into account the distinction between negative and positive shock, rather it is mechanically defining it as we did. Although we believe that such definition causes no loss of generality. Furthermore we have taken such decision after that we had observed the meaningless role of $\beta$ in the positive shock. Hence the decision of getting rid of it seemed harmless to us.

An other detail that could leave some readers unsatisfied is the decision of measuring only the euclidean distance between the simulated average price series, and the experimental series. One could obviously wonder why we did not measured the sum of the least squares, and run some
Figure 11: Average distance between PW’s data and our replication with negative shock.

Figure 12: Average distance between PW’s data and our replication with positive shock.
econometric analysis. Such decision has two main roots. First of all, you may find the standard deviation between simulated series and the experimental one in the Appendix, since we decided to implement them for sake of completeness, even though, as you will see, it adds nothing really new. So the first reason is merely practical: the value of $\omega$ that minimizes the standard deviation between the two series is quite similar to the ones we previously identified. But then one may wonder why we did not use the sum of least squares. And here we come to the second reason. We are aware that our method could suffer some loss of generality. In economic literature whenever you need to measure a distance between two empirical series, the most natural way is to run some OLS. But yet, we wanted to focus on an other issue of this topic.

Let’s go back to the start, and recall for a moment what we have done so far. We have taken a purely experimental work that showed evidence of an hypothetical phenomenon, identified by the authors by money illusion. Then we have taken such experimental data, and we have tried to come out with a theoretical model which was able to replicate such results, and give them a sort of “theoretical foundation”. We shall not forget that money illusion has no microfoundation up to now. The authors simply recorded the sign of a certain phenomenon; they deeply analysed and described the data, but were not able to give them a model able to legitimate such results. And now we made a tiny step further: we have made an attempt in the direction of finding a model for money illusion.

This is the reason why we simply left the distance between the two series. We wanted to show that our replication is not that bad after all. The error you see in the distance is actually how close we got in replicating the results coming from the behaviour of real human beings. We did provide a model which is in principle able to explain, and reproduce, the actions taken by someone in a lab somewhere in the world. One should read our results and appreciate how close we got in absolute terms. There is no real methodological thought behind this: we only believed that it would have been easier to interpret and to appreciate.

Finally, despite of the simpleness of our model, we are softly proud of it, because we believe it succeeded in what an economic model is supposed to do: it helped in explaining a phenomenon of the real world. An economic model may be considered as a mental experiment. You imagine a simplified version of the reality, which is defined through some assumptions which need to have some links to the real world. The stronger these links are, the more likely it is that what happens in the experiment may happen as well in the real world, in a way or an other.

Finally we believe that a further step should be trying to really bring the money shock inside the model: there is room for a proper model of money illusion with money assumed as an endogenous variable also in a model like this.
A Appendix

A.1 Payoff Functions and Strategic Complementarity

The payoff function used throughout every treatment is the following, and everything is evaluated at time $t$. Therefore there is no memory in the computation of the profits.

$$
\pi_i(P_i, \bar{P}_{-i}, M) = \frac{V \cdot \left[ \frac{1 + a \cdot \Delta^2}{1 + b \cdot \Delta^2} \right]}{1 + c \cdot \left[ \left( \frac{P_i}{M} - \frac{P_s}{M} \right) - d \cdot \Delta + e \cdot \arctan(f \cdot \Delta) \right]^2},
$$

where $\Delta$ is

$$
\Delta = \frac{\bar{P}_{-i} - P_s}{M}.
$$

$M$ represents the supply of money. During the experiment it will shift from $M_0$ to $M_1$, according to the money shock. $P_i$ represents the chosen price by the player, and $\bar{P}_{-i}$ represents the average price of $n - 1$ group members. See that $P_i$ does not go into the computation of the average price, in order to avoid complex calculations for the players. The sequence $a, b, c, d, e, f$ and $V$ represents a series of numerical parameters, with no particular economic meaning. They were the same in all the treatments and were given by $a = 0.5$, $b = 0.6$, $c = 27$, $d = 1$, $e = 0.05$, $f = 20$ and $V = 40$.

A.2 Several Changes to Our Model

Here we display several twists that we implemented to our model. First we show in detail which are the patterns caused by different values taken by omega, since the 3d-plot may look overflooded of information at a first glance.

In Figure 13 and 14 instead we show the results with the same values of $\omega$, but we have tested two cases. In the former plot we implemented a little shock, going from an initial money supply $M_0 = 42$, to $M_1 = 35$. In the latter we stressed the effects of the shock, enlarging money distance: from $M_0 = 42$ to $M_1 = 7$. In Figure 15 you can see what we mentioned in Section 3.2, that is the misalignment between the experimental and simulated series with the value of $\omega$ equal to 0.2, that is the value that gave a good fitness in the negative shock case.

A.3 Standard Deviation from Experimental Series

We show the plots of the standard deviation of the average simulated price with the experimental one, and we can see that there are few differences with the euclidean distance case. In Figure 16 and 17 you can see the FT squared distance.

In Figure 18 and 19 instead you can see the squared distance from PW’s data.
Figure 14: “Large shock”, going from $M_0 = 42$ to $M_1 = 7$.

Figure 15: FT’s data about positive shock, compared to our replication with previous $\omega = 0.2$.

Figure 16: Standard deviation between FT’s data and our replication: negative shock.
Figure 17: Standard deviation between FT’s data and our replication: positive shock.

Figure 18: Standard deviation between PW’s data and our replication: negative shock.
A.4 The Role of the Switch Parameter

One could wonder about several shortcomings of our work, though probably the most salient one is about the switch parameter $\beta$, which has been defined in an arbitrary way. But we would like to give an explanation about our choice.

When we started writing the first equations our aim was to avoid any possible loss of generality. Therefore we thought of a parameter that could drive the magnitude of the shock, but we immediately acknowledged that it was pointless in the negative shock case, since it did not add anything useful to the model. The fit was good enough to avoid us troubling with an other parameter, which we would have to give it an explanation anyway.

But then we saw that our model caused a little undershooting in the achievement of post shock equilibrium in the positive case. The logarithm has the nice feature of disappearing whenever its argument is equals one, that is in our case when money is kept constant. However, when the ratio between current money and previous money was smaller than one, we would incur in a negative quantity, which had the undesired effect of causing a little undershooting in our adjustment process. The pattern still remained really close to the experimental data, though such undershooting was really disappointing. Therefore we came up with the idea of simply decreasing $\beta$, which was simply equal to one in the negative case, having no magnitude effect. Then we kept decreasing it, and we saw that the smaller it was, the better was the fit in the case of a positive shock between our replication, and FT’s results. And we interpreted it as the fact that the bias $\theta_t$ played really no role in the case of a positive shock. Rather than being a shortcoming such definition of $\beta$, we believe it is a natural consequence of the asymmetric effect of money shocks.

Of course, our results between negative and positive case are asymmetrical also because of the eventual presence of $\beta$, but the results we found so far did not differ in the trend of the dynamics; there was only a misalignment in the last periods of the dynamics, as Figure 20 shows for $\beta = 1$.

Hence we take this sort of convergence process displayed in Figure 20 as a legitimacy of our definition of $\beta$ in [9].

Figure 19: Standard deviation between PW’s data and our replication: positive shock.
Figure 20: Convergence of the simulated series to the experimental series as $\beta$ decreases, until it disappears.
References


