Reducing the gender gap:

Biases in understanding delays in personnel policies

Inge Bleijenbergh, Jac Vennix, Eric Jacobs, Marloes van Engen

Radboud University Nijmegen, April 1, 2011

Reducing the gender gap:

Biases in understanding delays in personnel policies

Abstract

This paper builds on previous research in stock-flow experiments. 168 Participants were asked to recommend a hiring percentage of female professors they thought necessary to reach gender balance in an unbalanced personnel situation and to estimate the timeframe needed to reach this. Additionally, participants were asked to write down a justification for their choices. These justifications provide insight into the reasoning people use in stock-flow tasks. The experiment confirmed that people have difficulties understanding the distinction between stock and flows, and in particular underestimate the effect of average delays. Moreover it showed that not only cognitive bias, but also political considerations play a role in reasoning on this task. Further research is needed to examine the interplay between these two.

Keywords: bathtub dynamics, cognitive bias, stock-flow tasks, delays, gender balance, personnel policies.

Introduction

Over the last decade a number of experiments have been conducted under the heading of bathtub dynamics also known as stock-flow experiments. A variety of tasks has been used to explore cognitive biases of people in decision making (e.g. Ossimitz, 2002; Kainz and Ossimitz, 2002; Booth Sweeny and Sterman, 2000; Sterman, 2002). The most simple and most frequently used are the bathtub, the cash flow and the department store task. These consist of one stock with an in- and outflow. More complex tasks are the

Manufacturing Case and the CO_2 zero emissions task (Booth Sweeny and Sterman, 2000; Sterman and Booth Sweeney, 2002). The first consists of one stock and in- and outflow plus a negative feedback loop and a delay. The second has two stocks with in- and outflow and delays.

Two types of experiments have been conducted with the previous tasks. The first type focuses on establishing the difficulties people have in understanding the dynamics of simple stock-flow structures. The stock-flow experiments clearly reveal the difficulties people have in understanding the dynamics of simple stock-flow structures. Even predicting changes in the level of stocks in simple stock-flow structures presents many participants with insurmountable difficulties. People seem to mix up stocks and flows, when estimating the values of the stock variable given the behavior of the flow (Booth Sweeny and Sterman, 2000; Sterman, 2010; Cronin, Gonzalez and Sterman, 2009). In a simple task as the department store task it is common to find that less than half of the participants are able perform the task correctly (e.g. Ossimitz, 2002; Sterman, 2002, 2010; Pala and Vennix, 2005). The difficulties seem to be consistent over different tasks, and under a variety of conditions, e.g. the way information is presented, the difficulty of the task, motivation and outcome feedback, explicit direction towards accumulation (cf. Cronin, Gonzalez and Sterman, 2009).

The second type of stock-flow experiments studied the effect of system dynamics training on participants' performance in decision making. In this case results are mixed. Kainz and Ossimitz (2002) find significant improvements after a 90 minutes crash course in system dynamics. However, in a series of three experiments Pala and Vennix (2005) found mixed results. In the first, using the department store task, they find a significant difference in the performance of the task between the experimental (who participated in an introductory system dynamics course), and the control (introductory course in research methodology) group. The experimental group outperformed the control group. In the second experiment, a pretest – posttest without control group, the manufacturing task was used. Here performance in the posttest task is significantly worse than in the pretest task. In the third experiment, a pretest –posttest without control group using the CO_2 zero emissions task, no significant improvement in predicting changes in CO_2 emission could be established. In a carefully controlled experiment Sterman (2010) found significant improvements on the department store task after an introductory system dynamics course, but still overall performance in the posttest is far from 'perfect'. Finally, Phuah (2010) compared two approaches to improve participants' performance in the department store task, i.e. a 'guided graphical approach' and a 'running total and reflection approach'. Results show that participants using one of both outperformed a control group, although only the difference between the group using the graphical integration approach and the control group was significant. It is not really clear, however, why the running total approach is not effective in improving performance.

Thus far most experiments have relied on numerical data that do not shed much light on the cognitive reasoning people use when making their estimations. Analyzing participants' reasoning during these tasks, would not only increase insight in cognitive biases that lead to specific outcomes but also indicate what particular aspects need to be addressed in training to improve decision-making skills.

This paper reports on an experiment in which participants were not only asked to make a (numerical) choice, but also to provide a written explanation for their estimation. These written explanations provide additional insights into people's reasoning processes when confronted with a stock-flow task. Inspired by the CO_2 zero emissions task we developed a task that is socially relevant. This allows us to discuss the consequences of cognitive biases in decision-making on managing the issue. We focus on decision making on gender target figures in personnel policies, which is a hot topic in the Netherlands (Charter Talent naar de Top, 2011). By showing the dynamic structure that underlies this issue, we want to contribute knowledge to support policy makers in addressing the issue. The remainder of this paper explains the task and case, the selection of participants, the procedure and the results.

Task and case

The ‘female professors task’ we developed differs from the earlier experiments in two ways, namely in substance and stock flow structure. First, it differs in terms of the substance of the task in that it focuses on a social relevant issue in personnel policies, i.e. the gender (in) equality in higher management. The second difference is about the stock and flow structure. Rather than focusing on the relationship between the behavior of a stock and a flow, or controlling a manufacturing problem, the female professors task asks participants how to balance two different stocks, given an initial unbalanced personnel situation.

The issue is socially relevant, because Dutch organizations use target figures to increase the number of women in top positions, and in leading positions at universities in particular (Charter Talent naar de Top, 2011). In the Netherlands, the number of female full professors is very low in comparison to a) other countries, and b) the number of female students, PhD’s and assistant and associate professors. With eleven percent female professors in 2008 the Netherlands clearly ranks below the European average of 19% (Gerritsen, Verdonk and Visser, 2009; Van den Brink, 2009; Van Engen, Bleijenbergh and Paauwe, 2008; Van Engen, Bleijenbergh and Vinkenburg, 2010). In contrast, the number of female university students has risen above 50% since 2006, while the number of female PhD’s has risen from 24% in 1995 to 42 % in 2008 (Merens and Hermans, 2009).

Starting from the assumption that in general women and men are equally skilled to perform full professor tasks, one would assume a percentage female professors that is similar to the percentage female students (50%). However, this is clearly not the case. Typically, the argument which is frequently used is that it is just a matter of time for this balance to be accomplished, the so-called pipeline hypothesis (Xie and Shauman, 2003). As any system dynamicist will know this might take much longer than most people would normally expect (see also Dudley, 2007). This experiment is a first attempt to find out to what extent people understand the dynamic structure underlying changes in personnel composition and the consequences for decision making on gender target figures. Thus our research question is: to what extent are people able to estimate the correct hiring percentage of women to reach gender balance and what are the justifications they use for their estimation? We presented participants with a task in which we told them that the executive board of a Dutch university has decided to remove the gender gap of their university and to set an equal number of male and female full professors as a target. Participants read a one page description that: in the current situation there were about 300 positions available of which 33 were filled by women and 267 by men; the number of positions was to remain unchanged for the years ahead; a full professor, on average, stays on the job for a period of 10 years; 50% of the present students and 45% of the present PhD's is female; the board had recently decided to introduce hiring quota to ascertain that an equal gender balance, specifically 50% of the professors, would be accomplished in the near future. Participants were put in the position of advisor to this board and were given two questions:

a) What annual percentage of female professors would you recommend to hire as of 2010, given that objective?
b) How much time do you think it takes to realize the goal of equal gender balance?

Next, they were asked to explain why they had chosen this percentage and target year (see Appendix 1 for task). The task took around fifteen minutes for participants to accomplish. The dynamic structure of the task can be modeled as in Figure 1.

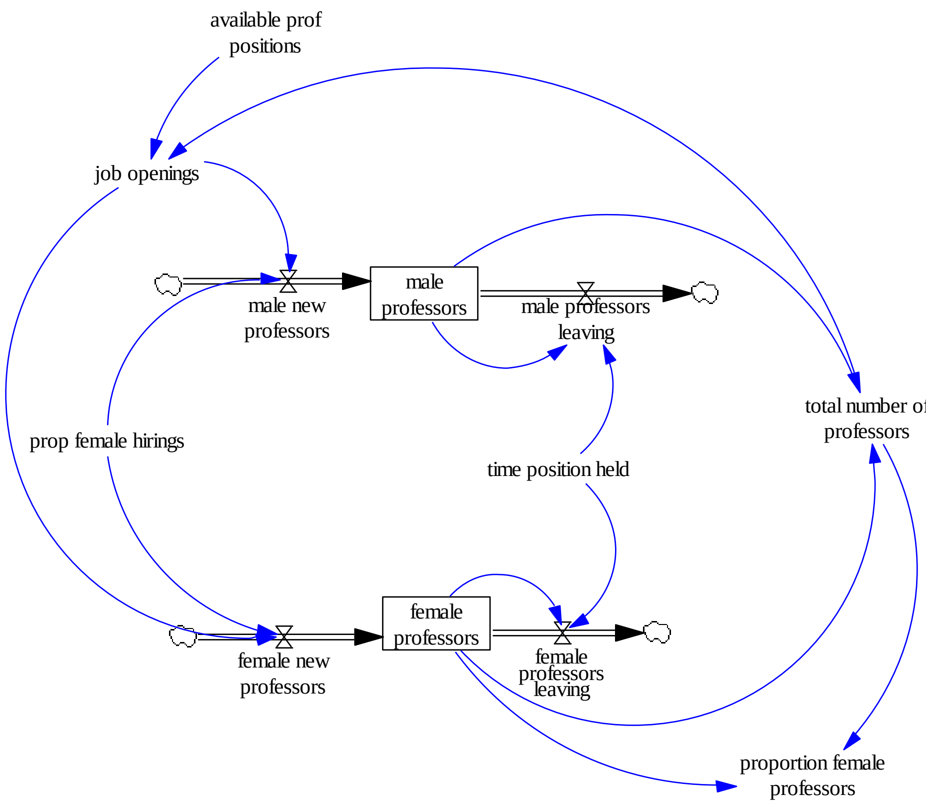

Figure 1: System dynamics structure of 'female professors task'

The stock and flow structure refers to a task where participants need to balance two different stocks, given an initial unbalanced situation. The first part of the task (recommending the percentage) is simple in the sense that, in order to reach the goal, the lower boundary of the hiring percentage of female professors is 50%. The second part of the task (i.e. estimating the time frame to reach an equal gender balance) is more complicated, because participants will have to take into account that the dynamics of this structure follows a growth path, which slows down over time. In our experiment for instance, female professors that are hired at a certain point in time also leave, slowing down the increase of the proportion of women faculty. As a result reaching the target will take longer than might be expected.

Participants and procedure

The experiment was conducted with three different groups of students at two different universities in the Netherlands. All students followed courses as part of a teaching program on Human Resource Studies or Human Resource Management. As students are following higher education and already finished introductory courses on HRM, we may consider them basically skilled for and interested in addressing social issues in the field of personnel policies.

The participants answered open questions in Dutch or English. Not all participants answered the questions completely. The information on sex, recommended percentage or year the 50% target is reached is missing for some participants. We decided to exclude those participants from further analyses. We also decided only to include the Dutch respondents in our sample due to small imperfections in translating the task in English. Of the original 195 participants 27 were excluded from further analyses. The most important characteristics of the remaining 168 participants are presented in Table 1.

Table 1 Characteristics for all participants and per subgroup

	Total (n=168)	**Subgroup 1 Autumn 2008 (n=51)**	**Subgroup 2 Spring 2009 (n=91)**	**Subgroup 3 Autumn 2009 (n=26)**
Age ($\bar{x}$)	22.0	22.2	21.6	22.9
Sex (%)				
Female	80.4	84.3	81.3	69.2
Male	19.6	15.7	18.7	30.8
Teaching program (%)				
HRS/HRM	80.4	96.1	72.5	76.9
Law&Management	9.5	0	12.1	19.2
Organization Studies	5.4	0	9.9	0
Other	4.8	3.9	5.5	3.8

The mean age of the total group of 168 participants is 22 years, while the average ages of subgroups 1, 2 and 3 are 22.2, 21.6 and 22.9, respectively. Of the total group around 80% is female. In group 1 this percentage is slightly higher (84.3%), while in group 3 it is lower (69.2%). Finally, as a consequence of conducting the experiment in courses on Human Recourse Management, the majority is Human Resource student. However, subgroup 2 and 3 contain also around 20% students engaged in related studies such as Law, Management and Organization studies.

Results

This section presents the estimations of the participants as well as their justifications. To analyze the estimations we discuss the frequencies of recommended percentage and estimated year, the correlation between recommended percentage and estimated year, and finally the discrepancy between estimated year and actual year (as calculated by the simulation model given a particular percentage). To analyze the participants justifications we performed content analysis to identify underlying patterns in reasoning.

First, we present the frequencies of recommended percentage of female appointments (further denoted as recommended percentage) and year that an equal gender balance is met (further denoted as target year). Kolmogorov-Smirnov's tests showed that the data for recommended percentage, $Z = 0.21$, $p < .001$, and for target year, $Z = 0.30$, $p < .001$, are not normally distributed. The results are presented in Table 2.

The median of the percentage participants recommended is 50%. The scores range from 1.5 till 100 and half of all the participants recommended a percentage in between 23.5% and 70%. The median of the target year is 2020. The scores range from 2012 till 2060 and half of the participants estimated gender balance to be reached between 2018 and 2025.

Table 2 Recommended percentage of female professors hired and estimated target year (n=168)

Recommended percentage	**Frequency**	**Percentage**	**Cumulative percentage**
0– 49	57	33.9	33.9
50 (exact)	50	29.8	63.7
51-100	61	36.3	100.0
Mdn	50%		
Min-Max	1.5-100		
Q1	23.5		
Q3	70		
Estimated target year	**Frequency**	**Percentage**	**Cumulative percentage**
2010-2019	60	35.7	35.7
2020 (exact)	60	35.7	71.4
2021-2060	48	28.6	100.0
Mdn	2020		
Min-Max	2012-2060		
Q1	2018		
Q3	2025		

Although an estimated target year before 2020 seems very optimistic, this can only be judged in relation to the recommended percentage. To evaluate the estimations we need to look at the correlation between recommended percentage and target year. Figure 3 presents a scatter plot with the recommended scores on estimated target year. The solid line shows the result of the simulation model.

The plot suggests four remarkable findings. First, the recommended percentages have a wide range but appear to be clustered around 50%. So participants have a tendency to choose a hiring percentage similar to the percentage of the target. Second, the target year is also spread and seems to be clustered around 2020. So, participants tend to choose a target year ten years ahead. Third, it seems that for the entire group of participants there is no correlation between recommended percentage and target year. Fourth, it shows that the majority of participants (75.7 %) give an estimation at the left hand side of the solid line. This indicates that in general participants underestimate what hiring percentage needs to be adopted to reach the target. To further analyze these findings we performed a series of quantitative analyses making a division between different groups of participants. We conducted the analyses for four different groups: 1. all participants; 2. participants who recommended less than 50%; 3. participants who recommended exactly 50%; 4. participants who recommended more than 50%. We selected these groups since they show different patterns in the scatter plot. The results of these analyses are presented in Table 3.

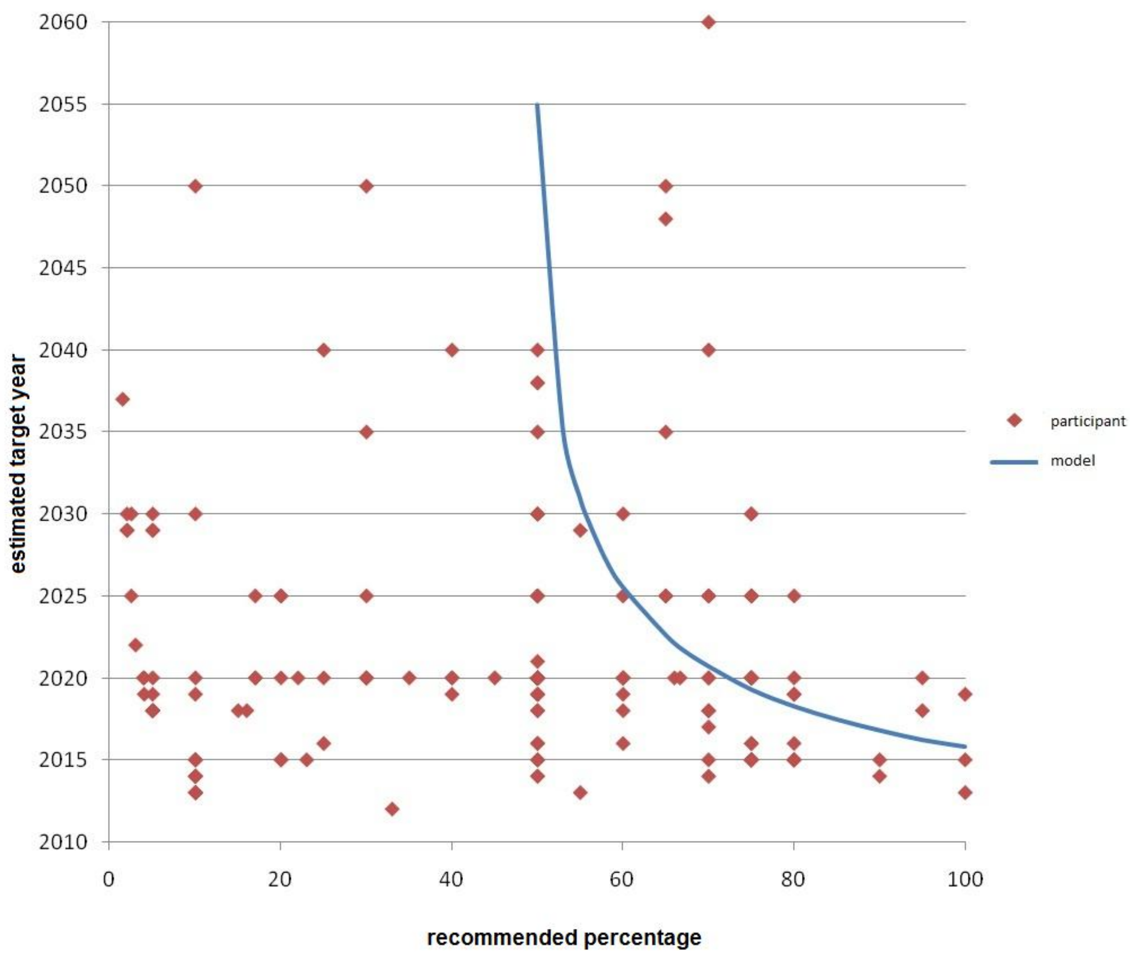

Figure 3: Scatter plot of recommended percentage by target year for all participants including a solid line representing the output of a simulation model

Table 3 Recommended percentage, target year, correlations between recommended percentage and target year, under and over estimation of target year for all participants and subgroups.

All participants (n=168):	
Recommended percentage (median)	50
Target year (median)	2020
Correlation Rec. percentage and target year[a]	-.088
Under/over estimation Target year (median)[b]	-6**
Participants recommending < 50% (n=57):	
Recommended percentage (median)	10
Target year (median)	2020
Correlation Rec. percentage and target year[a]	-.046
Under/over estimation Target year (median)	n.a.[c]
Participants recommending 50% (n=50):	
Recommended percentage (median)	50
Target year (median)	2020
Correlation Rec. percentage and target year[a]	n.a.
Under/over estimation Target year (median)[b]	-34**
Participants recommending > 50% (n=61):	
Recommended percentage (median)	75
Target year (median)	2020
Correlation Rec. percentage and target year[a]	-.391**
Under/over estimation Target year (median)[b]	-1

[a] Spearman's rank correlation (one sided test)
[b] Sign test (two sided test)
[c] not applicable
* p < .05, ** p< .01 (one sided tests)

The results for the entire group of participants show that the median of the recommended percentage and target year is 50 and 2020 respectively. This supports our first two findings from Figure 3. The correlation between recommended percentage and target year is negative, but not significant (r_s=-0.09). So, the absence of a correlation, our third finding, is also supported. Finally, we analyzed the extent to which participants over- or underestimated the target year given their recommended percentage. We could only calculate that for the participants who recommended a percentage of 50% or more because, obviously, with a percentage less than 50% one will never reach the target. In general, these participants appear to underestimate the target year with six years.

The results for the participants with a recommended percentage below 50% show that the median of the recommended percentage is 10 and of the target year is 2020. The correlation between recommended percentage and target year is negative, but not significant (r_s=-0.05). Finally, the extent in which participants over- or underestimated the target year given their recommended percentage is not applicable, as we could only calculate this for the participants who recommended a percentage of 50% or more, as stated above.

When looking at the participants who recommended exactly 50%, the results show that the median of the recommended percentage is 50, obviously, and of the target year is 2020. The correlation between recommended percentage and target year could not be calculated for the recommended percentage has only one value. Regarding the extent in which participants over- or underestimated the target year given their recommended percentage, the median is -34. So, in general this group appears to underestimate the target year considerably.

The results for the group of participants that recommended more than 50% show that the median of the recommended percentage is 75 and of the target year is 2020. The correlation between recommended percentage and target year is significant, negative and fairly strong (r_s=-0.39). So, the absence of a correlation in the entire group, our third finding, does not hold for this group of participants. When looking at the extent in which participants over- or underestimated the target year given their recommended percentage, the median is -1 and not significant. This means that in general, participants recommending more than 50% appear not to under- or overestimate the target year.

Justification: participants' explanations

To reveal patterns in the justification of the estimation, we performed content analysis. The written material was literally transcribed and coded with open codes by two researchers. We used inductive coding (Strauss and Corbin, 2008), to ensure that themes that we did not expect beforehand would be included. In deliberation with a third researcher, five categories of justifications were identified: i.e. equal treatment of men and women, the quality of applicants, women's (lack of) career motivation, the need to compensate for present inequalities, and numerical explanations. Thematic coding was checked by the other two researchers, until consensus about the coding was reached.

Interestingly, when it comes to the motivation of their advice most participants (115 or 68%) use an argument based on political considerations (i.e. the first four themes). Only 53 (32%) of the participants present a numerical explanation. We assume this majority of 'politically loaded' arguments has to do with the subject of gender target figures being a political sensitive issue in the Netherlands This results is probably the consequence of using a task on a social relevant issue. However, we wonder if it negatively affects the performance of the task. In the remainder of the paper we will first compare the estimations of the group that used numerical and the group that used 'non numerical' explanations, to check for eventual differences in performance. We present the results of these analyses in Table 4. Next, we analyze the numerical explanations s to understand the cognitive bias in decision making surrounding the issue.

Table 4 Recommended percentage, target year, correlations between recommended percentage and target year, under and over estimation of target year for participants who gave a non-numerical and a numerical explanation.

Participants with non-numerical explanations (n=115):	
Recommended percentage (median)	50
Target year (median)	2020
Correlation Rec. percentage and target year[a]	.010
Under/over estimation Target year (median)[bc]	-5**
Participants with numerical explanations (n=53):	
Recommended percentage (median)	50
Target year (median)	2020
Correlation Rec. percentage and target year[a]	-.382**
Under/over estimation Target year (median)[bd]	-34**

[a] Spearman's rank correlation (one sided test)
[b] Sign test (two sided test)
[c] n=72
[d] n=39
* p < .05, ** p< .01 (one sided tests)

Participants with non-numerical and numerical explanations.

When looking at participants using a non-numerical explanation for their estimations, the results show that the median of the recommended percentage is 50 and of the target year is 2020. The correlation between recommended percentage and target year is not significant (r_s=0.01). Regarding the extent in which participants over- or underestimated the target year given their recommended percentage, this could only be calculated for the participants who recommended a percentage of 50% or more. In general, participants who provided a non-numerical explanation and recommended a percentage of 50% or more appear to underestimate the target year with five years.

Finally, the results for participants providing a numerical explanation for their estimations show that the median of the recommended percentage is 50 and of the target year is 2020. The correlation between recommended percentage and target year is significant, negative and reasonably strong (r_s=-0.38). Regarding the extent in which participants over- or underestimated the target year given their recommended percentage, the median is -34. In general, the participants who provided a numerical explanation and recommended a percentage of 50% or more appear to underestimate the target year considerably. It appears that the underestimation of these participants is considerably larger than the underestimation of participants that recommended a percentage of 50% or more, while using non-numerical explanations.

Different results are found for insight in the negative relation between recommended percentage and target year. Although the underestimation is large, participants using numerical explanations seem to have better understanding of this relation. This result seems to indicate that making an effort to calculate helps the performance in this task, at least regarding understanding the negative relation between recommended percentage and target year. However, it does not seem to help predicting the target year. In the next section we elaborate on the justification of the recommendations of the participants who gave numerical explanations. To structure the analysis we identify patterns in reasoning for the groups that estimated a percentage below 50%, of 50% and above 50%.

Justification of the group recommending percentages below 50%

Of the participants using numerical explanations, 24.6% recommends a percentage below 50 percent, (compared to 33.9% of all participants). At first sight, this looks quite puzzling since it is obvious that with such percentages it is impossible to reach the target and the percentage of women full professors would even drop rather than increase with any percentage lower than 11 percent. A closer look at the explanations reveals that these participants use the following type of reasoning: the percentage is now 11%; it needs to go up to 50%; an increase of x% per year is thus required; the number of years * the percentage + initial value produces 50%. For example: "5% per year is feasible, which means raising the level of female professors with half of them in a period of eight years". (participant 11, estimated year 2018). Or: "Now 10% female professors. In 10 years a complete new group; in this way nobody will be laid off. Having an increase of 5% per year will result in 50%." (participant 194, estimated year 2019). Even for the person who selected 1.5% this reasoning holds, because this participants' target year is 2037 (participant 110), meaning 27 years * 1.5% = 40.5% + 11 = 51.5%. The reasoning seems logical: if the percentage is now 11% and increases by 5% per year, then after eight years the percentage will have passed 50%. However, the task requires participants to give an annual percentage of female professor to be hired (flow) and not the increase in percentage of the comparison between the two stocks. So they did not perform the task correctly.

Justification of the group recommending exactly 50%

Of the participants who used numerical explanations 45.3% recommended hiring 50% female professors (compared to 29.8% of all participants). These participants seem to understand the difference between stocks and flows, mostly arguing that if one uses a quota of 50% one must eventually end up with 50% female professors. However, as the scatter plot of Figure 3 indicates, all participants in this group underestimate the time which is needed to accomplish that target. Again the reasoning follows a 'static' pattern. Participants seem to use a discrete pipeline model: they assume that all professors will leave after exactly ten years. As a result the majority of these participants argue that since in ten years all available positions will be taken by new people, an annual allocation of half of the vacant positions to women, will result in realization of the desired situation in ten years from now (i.e. around 2020). For example: "300 positions. 10 years. 30 vacancies annually = 15 women * 10 years = 150. This is half of the total number of positions." (participant 188, obviously ignoring the initial situation). At first sight, this reasoning seems plausible. However, the task mentions that a full professor position is held for *on average* ten years. This would mean that, for example, some full professors would leave after two years and others only after twenty years. So, the task assumes an exponential rather than a discrete delay, which has consequences for the dynamic structure (Compare Grössler & Zock, 2010). If a policy of hiring 50% female professors is adopted, and a professor stays on the job for on average ten years, the percentage of female full professors will increase rapidly in the beginning, but the growth will slow down over time since also more female professors will be leaving.

Justification of the group recommending percentages above 50%

Of the group using numerical explanations, 28.3% selected a percentage above 50% (compared to 36.3% of all participants). The justification in the group,that recommends a hiring percentage above 50%, follows two patterns. Part of this group uses the same kind of reasoning as the previous group, but with a higher percentage and a corresponding shorter time frame, e.g.: "On average 10 year in a position. So about 1/10 of all professors quits every year (30), for which new ones need to be found. 60% of 30 = 18 women. Target is 150 women. (150-33) / 18 = 6.5 years." (participant 184, selecting 60% and 2016).

Other participants in this group seem to sense that recommending 50% will take a long time to accomplish an equal gender balance and opt for a higher hiring percentage, e.g.: "50% will make things equal in the very long run. 80% will make sure that things will be made equal in a much shorter period." (participant 38, mentioning as the target year 2015, still underestimating reaching the gender balance by 3-4 years). Two participants select a percentage of 100%, one of which estimates the gender balance reached by 2015 (participant 130) and the other by 2019 (participant 32). Although their intuition is correct that hiring 50% per year will take quite some time, none of these participants in this category makes a clear calculation. So, high percentages seem to be justified by an intuition about the long period it takes to reach equal gender balance rather than by a correct calculation.

Conclusions

In this paper we discussed a new experimental stock-flow task, 'the female professor's task', in which participants are asked to (1) recommend a hiring percentage for female professors, (2) estimate the year that a gender balance is reached, and provide a written justification for their choice. The data gave us insight in cognitive biases that hinder participants to solve the task correctly. Moreover, it shows that people have a tendency to justify their decisions with a 'political loaded argument' rather than a numerical explanation.

First, taking the results for the 168 participants together we found that both the recommended percentage and the target year have wide ranges: from 1.5 to 100% and from 2012 to 2060 respectively. This means there is a wide variety of recommended percentages and estimations of the target year. Second, in general there is a severe underestimation of the time needed to accomplish the target. We found that 75.7% of the participants estimated a target year that is lower than the year an equal gender balance would be reached considering the percentage of female professors they recommended to hire. Moreover, 33.9% of the participants even recommend a percentage by which an equal gender balance will be never reached. Participants seem to have difficulties to understand the effects of a certain hiring policy on the target set.

Third, there is no correlation between recommended percentage and target year overall. In general the participants do not show to understand that gender balance will be reached sooner when the percentage of female professors is higher. In contrast, for those who recommend a percentage of 50 or more there is a considerable correlation between recommended percentage and target year. It seems that those who recommend a higher percentage have a better understanding of the relationship between the stocks and flows. Fourth, a majority of 68% of the participants used political considerations rather than a numerical explanation to provide a justification of their estimation. The participants who gave a numerical explanation performed better when it comes to the relationship between recommended percentage and target year. However these respondents underestimate the percentage of women that should be hired to reach the target year considerably.

The fifth result concerns the justifications for the different estimations. Analysis of the numerical explanations shows that advising a hiring percentage below 50% is justified in terms of the percentage by which the stock should increase rather than the flow. These participants understand how fast the stock should increase, but do not succeed in translating this in a flow percentage. Recommendations of exactly 50% (and also above 50%) are combined with explanations relating to a discrete delay rather than an exponential delay. Finally, recommending a percentage above 50% is justified by an intuition that it will take long to reach gender balance rather than by a correct calculation based on an exponential delay.

Discussion

This study shows that participants not only have difficulties to understand changes in stocks and flows in general, but that they have a tendency to underestimate the time which is needed to balance two different stocks, given an initial unbalanced situation. By using the social relevant issue of gender target figures, we show that decision-makers need to be aware of this potential underestimation. Cognitive biases may eventually lead to a decision to hire a percentage of female candidates below the 50 percent that is needed to reach gender balance.

More in particular, it seems that participants have difficulties understanding the concept of average delay time. Overall, a large proportion of the participants estimate that the target will be reached in ten years. This points towards a mental model which uses a discrete delay or 'first in first out' (FIFO) approach. This is understandable as it requires much less mental effort to estimate the effects of a discrete delay than an exponential delay. However, in real world the process of workforce planning has the character of exponential delays (Grössler & Zock, 2010). Not understanding this phenomenon has consequences for personnel policies regarding hiring underrepresented groups such as women or migrants. Since we performed our experiment with students following courses on human resource management, we wonder if human resource managers may also underestimate the hiring percentage and time needed to reach an equal gender balance. This could be a topic of a follow up study.

Moreover, the experiment showed that with such a sensitive issue the majority of participants (68%) use 'politically loaded' rather than numerical explanations to justify their choice. The participants using numerical explanations on average were aware that an equal gender balance would be reached earlier once the percentage female professors hired would be higher. A follow up study with a more 'neutral task' would be needed to examine to what extent the underestimation is caused by the political sensitivity of the issue and to what extent by cognitive bias.

We suggest that HRM students should be educated in the difference between stocks and flow and the delay processes in personnel policies. Given the fact that the experimental group is being educated for positions in HRM this suggests that training in system dynamics to support advising on personnel policies is beneficial. We support the statement by Cronin et al. (2009; 129) that we need to improve our ability to understand and manage complex systems affecting organizations and society. Currently, experiments are underway to better understand the difficulties people have in understanding the concept of average delays. To conclude, we would like to point out that so far stock-flow experiments have focused on a variety of tasks, without a systematic research program delineating the core tasks and difficulties people have in working with elements of complex systems. Such a research program could help tremendously to focus research efforts, and to build a coherent theory and set of experiments.

References

Booth Sweeney L. Sterman JD. 2000. Bathtub dynamics: initial results of a systems thinking inventory. *System Dynamics Review,* **16**(4): 249-286

Cronin, M.A., Gonzalez, C., Sterman, J.D. 2009. Why don't well-educated adults understand accumulation? A challenge to researchers, educators and citizens. *Organizational Behavior and Human Decision Processes*, (108): 116-130.

Charter Talent naar de Top (2011). Ondertekenaars, http://www.talentnaardetop.nl/Charter/Preambule.htm, viewed at Januari 20 2011.

Dudley, R. (2007) The Equity Supply Chain: Is it the Cause of So Few Women in Management and Leadership Positions? *Proceedings of the 27th International Conference of the System Dynamics Society*, Boston, MA, USA http://www.systemdynamics.org/conferences/2007/proceed/papers/DUDLE171.pdf.

Gerritsen, M, Verdonk, T., Visser, A. 2009. *Monitor Vrouwelijke Hoogleraren 2009, Stand van zaken vrouwen in universitaire functies en wetenschappelijke besluitvormingsorgansen.* Viewed at http://www.sofokles.nl/downloads/actueel/Monitor.28.09.2009.pdf

Größler, A., Zock, A. 2010. Supporting Long-term Workforce Planning with a Dynamic Aging Chain Model – a case study from the service industry. *Human Resource Management* 49(5): 829–848.

Kainz D. Ossimitz G. 2002. Can students learn stock–flow-thinking? An empirical investigation. *Proceedings of the 20th International Conference of the*

System Dynamics Society, Palermo, Italy. System Dynamics Society: Albany, NY.

Merens, A., Hermans, B. 2009. *Emancipatiemonitor 2008.* Sociaal Cultureel Planbureau & Centraal Bureau voor de Statistiek: Den Haag.

Ossimitz G. 2002. Stock–flow-thinking and reading stock–flow-related graphs: an empirical investigation in dynamic thinking abilities. *Proceedings of the 20th International Conference of the System Dynamics Society*, Palermo, Italy. System Dynamics Society: Albany, NY.

Pala, Ö, Vennix, J.A.M. 2005. Effect of system dynamics education on systems thinking inventory task performance. *System Dynamics Review*, **21**(2): 147-172.

Phuah, T.L.K. 2010. Can people learn behaviours of stocks and flow using their ability to calculate running total? An experimental study. *Proceedings of the 28th International System Dynamics conference*, Seoul, Korea System Dynamics Society: Albany, NY.

Sterman, J.D. 2002. All models are wrong: reflections on becoming a systems scientist. *System Dynamics Review* **18**(4): 501–531.

Sterman, J.D. 2010. Does formal system dynamics training improve people's understanding of accumulation? *System Dynamics Review* **26**(4): 316–334.

Sterman JD, Booth Sweeney L. 2002. Cloudy skies: assessing public understanding of global warming. *System Dynamics Review* **18**(2): 207–240

Strauss, A, Corbin, J. 2008 *Basics of qualitative research; techniques and procedures for developing grounded theory*, Sage Publications: Thousand Oaks.

Van den Brink, M. 2009. Behind the scenes of science; gender practices in the recruitment and selection of professors in the Netherlands, Ipskamp drukkers: Nijmegen.

Van Engen, M., Bleijenbergh, I., Paauwe, J. 2008. *Vrouwen in hogere wetenschappelijke posities aan de Universiteit van Tilburg: Processen van instroom, doorstroom en uitstroom nader bekeken.* Universiteit van Tilburg: Tilburg
http://www.uvt.nl/medewerkers/vrouwen/onderzoek.pdf

Van Engen, M.L, Bleijenbergh, I.L., Vinkenburg, C.J. 2010. *Structurele en culturele belemmeringen in de doorstroom van vrouwen naar hogere functies binnen de TU Delft; Eindrapport Onderzoeksproject Talent naar de Top* Technische Universiteit Delft: Delft.

Xie, Y. and Shauman, K. (2003) *Women in Science: Career Processes and Outcomes*. Harvard University Press, Cambridge, MA.

Appendix 1

Experiment appointing of female professors

The Board of xxx[1] University has decided to drastically increase the number of female full professors. The target is to have 50% female full professors at the university in the long run. Currently, the percentage of female full professors is 11%. For 15 years already, the percentage of female students is around 50% while 45% of the current PhD students are female. It is expected that this percentage will increase in the years to come. This situation has lead to the target being set at 50% for female full professors. Considering the financial situation at institutions for higher education the assumption is that the number of full professor positions will not increase. The change therefore has to occur within the current number of full professor positions.

At xxx University the number of positions for full professors is about 300. As stated, it is expected that this number will not increase in the near future. There are about 33 female and 267 male full professors. On average, a full professor position is held for 10 years. After this period, the professor retires or accepts a position at another university, so a new professor can be appointed at that position. As mentioned, the target is to have 50% female full professors at the university.

Within the Board there has been a long discussion about the strategy that has to be pursued to achieve this goal and the Board is convinced they have to implement a quota. This means that faculties will be forced, as of 2010, to annually appoint a certain percentage of women on vacant professor positions.

[1] Name masked for review purposes

The Board asks your advice about how high this percentage should be and when you think the target (of 50% female and 50% male professors) will be achieved.

My advice on annual percentage of female appointments:……...%

Using this percentage the target will be achieved by the year 20....

Could you please briefly justify your choices below?

Finally we would like to ask you a couple of additional questions.

Gender: man/woman
Year of birth: 19..
Study:

Thanks for your participation!

Appendix 2: equation listing of female professors model

(01) available prof positions=
    300
  Units: **undefined**

(02) female new professors=
    job openings*prop female hirings
  Units: person/Month

(03) female professors= INTEG (
    female new professors-female professors leaving,
      33)
  Units: person

(04) female professors leaving=
    female professors/time position held
  Units: person/Month

(05) FINAL TIME = 1000
  Units: Month
  The final time for the simulation.

(06) INITIAL TIME = 0
  Units: Month
  The initial time for the simulation.

(07) job openings=
    available prof positions-total number of professors
  Units: job

(08) male new professors=
    job openings*prop female hirings
  Units: person/Month

(09) male professors= INTEG (
    male new professors-male professors leaving,
      267)
  Units: person

(10) male professors leaving=
    male professors/time position held
  Units: person/Month

(11) prop female hirings=
    0.5
  Units: Dmnl

(12) proportion female professors=
female professors/total number of professors
Units: Dmnl

(13) SAVEPER =
TIME STEP
Units: Month [0,?]
The frequency with which output is stored.

(14) time position held=
120
Units: Month

(15) TIME STEP = 1
Units: Month [0,?]
The time step for the simulation.

(16) total number of professors=
female professors+male professors
Units: person