

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/160532>

Please be advised that this information was generated on 2018-10-16 and may be subject to change.

Understanding decision making about balancing two stocks: the faculty gender balancing task

Inge Bleijenbergh,^{a*} Jac Vennix, Eric Jacobs^a and Marloes van Engen^b

Abstract

This paper contributes to the understanding of dynamic decision making by exploring how people understand the dynamic system underlying a task with modest dynamic complexity. Using the faculty gender balancing task, we experimentally test the extent to which participants correctly estimate the inflow needed to balance two unbalanced stocks and how they substantiate their decision. The results, based on performance data for 133 participants, show that people have difficulty determining the correct hiring percentage. Three-quarters of the participants recommended a hiring percentage that was too low to balance the stocks in time or to balance them at all. Analysis of the substantiations showed that nearly half of the participants underpinned their decision with calculations, while 45 percent used intuitive estimations. The calculations seem to be based on three heuristics, reasoning patterns on which participants erroneously relied to assess dynamic behavior. Contrary to what has been found so far, the participants who used intuitive estimations to support their decisions performed relatively better on the task. © 2016 The Authors. *System Dynamics Review* published by John Wiley & Sons Ltd on behalf of System Dynamics Society

Syst. Dyn. Rev. (2016)

Additional supporting information may be found in the online version of this article at the publisher's web site.

Introduction

Over the last decades, there has been a consensus that individuals need system thinking skills to make effective decisions in dynamic environments (Booth Sweeney and Sterman, 2000; Gonzalez and Wong, 2012). These system thinking skills include the ability to understand the dynamic behavior of a problem over time, to discover the feedback processes that play a role, to identify the stock–flow relationships that are involved, to recognize time delays and their effect and to identify nonlinearities (Booth Sweeney and Sterman, 2000, p. 250). A number of experiments have been conducted under the heading of bathtub dynamics, also known as stock–flow tasks, to explore the difficulties people have in decision making when facing dynamic problems (e.g. Booth Sweeney and Sterman, 2000; Gonzalez and Wong, 2012; Kainz and Ossimitz, 2002; Kampmann and Sterman, 2014; Ossimitz, 2002; Pala and Vennix, 2005; Sterman, 2002). The simplest and most frequently used tasks in these studies are the bathtub task, the cash flow task and the department store task. These tasks consist of decision making about one stock with inflow and outflow

^a Institute for Management Research, Radboud University Nijmegen, HK Nijmegen, The Netherlands

^b Department of Human Resource Studies, Tilburg University, LE Tilburg, The Netherlands

* Correspondence to: Inge Bleijenbergh. E-mail: i.bleijenbergh@fm.ru.nl

Accepted by Markus Schwaninger, Received 17 September 2014; Revised 11 April 2016; Accepted 27 April 2016

System Dynamics Review
System Dynamics Review (2016)
Published online in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/sdr.1548

© 2016 The Authors. *System Dynamics Review* published by John Wiley & Sons Ltd on behalf of System Dynamics Society

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

without any feedback processes, delay or nonlinearity. Tasks with modest levels of dynamic complexity include the manufacturing task and the CO₂ zero emissions task (Booth Sweeney and Sterman, 2000; Sterman and Booth Sweeney, 2002). The manufacturing task consists of decision making about one stock with inflow and outflow plus a negative feedback loop and a delay (Booth Sweeney and Sterman, 2000, p. 272), whereas the zero emissions task includes two stocks with inflow and outflow and a delay (Sterman and Booth Sweeney, 2007, p. 215). More recently, the faculty gender balancing task, which is also a task with modest dynamic complexity because it includes two stocks with inflow and outflow, negative feedback loops and a delay (Table 1), was introduced (Bleijenbergh *et al.*, 2011). In this paper, we contribute to providing a better understanding of decision making about tasks with modest dynamic complexity by exploring the difficulties people have in decision making about this last task: the faculty gender balancing task.

Experimental research has established the difficulties people have in understanding the dynamics of even simple stock–flow structures (e.g. Ossimitz, 2002; Sterman, 2002, 2010; Pala and Vennix, 2005). Insight into these difficulties has helped to identify heuristics, which we define as reasoning patterns on which people erroneously rely to assess the behavior of dynamic systems of modest dynamic complexity (cf. Sterman, 2010, p. 328). Based on Sterman and Booth Sweeney (2002, 2007), we assume that heuristics, despite being false representations of dynamic behavior, may be helpful in supporting decision making in everyday tasks with low dynamic complexity, such as filling a bath tub. If the delay is short, if people get timely feedback on the outcome, if opportunities for corrective action are frequent, and if costs of errors are small, using reasoning patterns that simplify the understanding of dynamic behavior is not problematic. However, with tasks of modest dynamic complexity, such as addressing CO₂ emission, the delay is substantial, feedback on the outcome is often delayed, opportunities for corrective action are small, and the social, environmental and economic costs of errors are large (Sterman and Booth Sweeney, 2007, p. 233). In such cases, the use of heuristics may need to be prevented. Thus identifying reasoning patterns on which people erroneously rely to understand the behavior of dynamic systems of modest dynamic complexity is important in understanding and improving dynamic decision making.

Table 1. Stock–flow tasks

Task	One stock	Two stocks	Feedback	Delay
Bath tub	x			
Cash flow	x			
Department store	x			
Manufacturing	x		x	x
CO ₂ zero emission		x		x
Faculty gender balancing		x	x	

An important heuristic identified, called the *pattern matching heuristic* (Sterman and Booth Sweeney, 2002) or *correlation heuristic* (Cronin *et al.*, 2009; Sterman, 2010), is the tendency to “erroneously assume that the behavior of a stock matches the patterns of its flows” (Cronin *et al.*, 2009, p. 116). The correlation heuristic shows how people fail to grasp that a stock rises only when the inflow exceeds the outflow and that a stock will fall only when the outflow is larger than the inflow (Sterman, 2010, p. 316). Experiments with decision tasks of more modest dynamic complexity have revealed additional difficulties that people have in understanding the dynamic behavior of dynamic systems. In stock–flow tasks with one stock, feedback and a delay, participants not only poorly understand stock and flow relationships but also underestimate time delays between changes in inflow and the stock level (Booth Sweeney and Sterman, 2000). In stock–flow tasks with two stocks and a delay, participants also underestimate the time delay. Moreover, people underestimate the inertia of the dynamic system in general (Sterman and Booth Sweeney, 2002, p. 233). The reasoning patterns underlying these additional difficulties in more complex decision tasks have not yet been identified, and this paper addresses this research gap.

The experiments reveal that the difficulties in understanding dynamic behavior appear in different tasks and under a variety of conditions, e.g. the difficulty of the task, the display of data, the context of the task, the motivation of the participants and the provision of feedback about the outcomes (cf. Cronin *et al.*, 2009). However, recent research suggests that the way information is presented in the task may increase individuals’ understanding of dynamic behavior. In particular, presenting a stock–flow problem by comparing it with other stock–flow problems with behavioral similarity increases individuals’ performance on the task (Gonzalez and Wong, 2012). Moreover, some evidence suggests that verbal presentation compared with graphical presentation may increase individuals’ understanding of a dynamic system (Fischer and Degen, 2012). The integration of a graphical presentation into a narrative also seems to lead to greater understanding (Phuah, 2010).

Thus far, most experiments have relied on numerical data that do not shed much light on the way participants reason when making decisions on dynamic problems. Korzilius *et al.* (2014) made a first effort to analyze the reasoning behind a simple stock and flow task and showed that participants used a variety of reasoning patterns. Kampmann and Sterman (2014) used post-test questionnaires to measure how participants understood a dynamic system. We further expand this line of research to stock and flow tasks of modest dynamic complexity. Inspired by the CO₂ zero emission task (Sterman and Booth Sweeney, 2002), we use a stock–flow task in the context of a real-world issue that allows us to discuss the public policy consequences of the difficulties people have in understanding dynamic problems. The faculty gender balancing task asks participants to make a (numerical) decision about the inflow needed to reach a target and to verbally substantiate this decision. In

contrast to the CO₂ zero emission task, which requires that participants estimate the effect of reducing one stock (greenhouse gas in the atmosphere) on another stock (global temperature), we ask participants to balance two stocks (the number of male full professors and the number of female full professors) in an initially unbalanced situation. Analyzing participants' reasoning patterns may indicate the difficulties that need to be addressed to improve decision making about complex dynamic problems. In summary, with this experiment, we aim to determine how people understand the dynamic system underlying balancing two stocks in an initially unbalanced situation and what the consequences are for decision making. Thus our research question is as follows: "To what extent do participants performing the faculty gender balancing task correctly estimate the hiring percentage needed to balance two stocks in an initially unbalanced situation, and how do they substantiate their decision?" In the remainder of this paper, we describe the task and case, the selection of participants, the procedure and the results, and we finish by presenting a conclusion and a discussion of the meaning of our results for understanding stock–flow failure.

Dynamic decision task

The "faculty gender balancing task" we developed differs from other stock–flow tasks in that it involves two stocks and two negative feedback loops. The "faculty gender balancing task" asks participants to balance two different stocks—male full professors and female full professors—to reach a balance between them (an equal proportion of female and male full professors). Balancing unbalanced stocks, such as the number of trained and untrained employees, is a common dynamic problem in personnel policies, and prior research has suggested that people underestimate time delays in this context (Gröbler and Zock, 2010).

Reaching an equal balance between stocks of male and female full professors is a real-world issue for organizations in industrialized countries such as the UK and the Netherlands. Many of these organizations use target figures to increase the number of women in higher management and in leading positions at universities in particular (Charter Talent to the Top, 2016). In the Netherlands, the number of female full professors is low in comparison to that in other countries and with the number of female students, PhDs and assistant and associate professors. With an average of 15 percent female full professors in 2011, the Netherlands clearly ranks below the European average of 19 percent (Gerritsen *et al.*, 2012; Van den Brink, 2009). In contrast, the number of female university students has been above 50 percent since 1999 (Merens *et al.*, 2012).

Starting from the assumption that women and men are generally equally skilled to perform full professor duties, one would assume a percentage of

female full professors that is similar to the percentage of female students (approximately 50 percent). However, this is clearly not the case. Typically, the argument 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). However, as system dynamicists argue, the pipeline delay hypothesis is not sufficient to explain the relatively small number of women in higher management. Dudley (2007) shows that, with the inflow proportions of women in basic careers during the prior seven decades, 8 percent more women should have been in senior positions in 2007 than real-world data report (2007, p. 5).

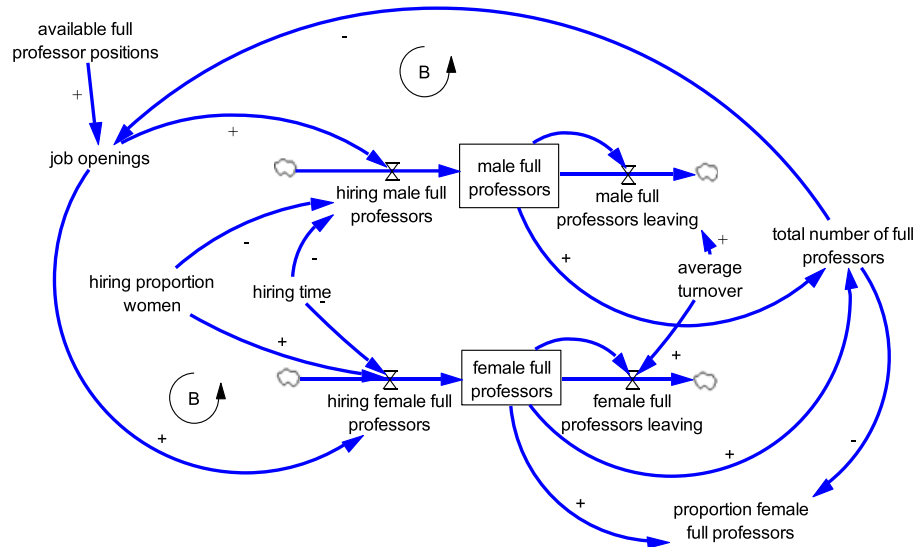
We presented participants with a task in which we told them that the board of a university aims to reach a gender balance in the number of full professors at their university in 15 years. In a one-page description, the participants read that the current situation is that there are 1000 full professor positions, of which 10 percent are occupied by women and 90 percent by men; the number of full professor positions will remain unchanged for the years to come; the turnover of full professors is 7 percent per year, for both men and women¹; 50 percent of the students were women in recent decades; and the board has decided to implement a hiring quota to achieve the goal of gender balance. The participants were put in the position of chair of the board and were asked to set the percentage of female full professors they would require the schools to hire to reach a gender balance in 15 years. In addition, the participants were told that a sufficient number of men and women were available to fill the positions, regardless of the percentage they select. Next, they were asked to briefly explain why the objective of equal gender balance would be achieved in 15 years, given the percentage they selected. (See Appendix 1 for a full description of the task.)

The faculty gender balancing task was adapted from a pilot study (Bleijenbergh *et al.*, 2011) in three ways. First, we used a fixed target of 15 years to reach a balance between the stocks rather than leaving the target year open. Second, we added more specific information about the yearly turnover: the level of turnover is the same for female and male full professors and is equally distributed over the different age groups. Third, we stated that the participants, in the role of chair of the board, were supposed to agree with the goal of reaching a 50:50 gender balance and were responsible for implementing it. These adaptations were made to facilitate comparisons of the answers, to ensure that the participants had all relevant information to make a correct estimation of the inflow percentage needed to balance the two unbalanced stocks at a given moment in time, and to prevent them from entering into moral discussions about gender quota, as happened in the pilot study.

The task can be modeled as a stock–flow structure, as presented in Fig. 1 and the supporting information can be found in the online version of the article. Participants need to balance two different stocks to reach an equal

¹A case study at a Dutch university suggested the average turnover of both male and female full professors is 15 years (Van Engen *et al.*, 2008).

Fig. 1. Structure of the system dynamics model in the “faculty gender balancing task”



proportion of female and male full professors given an initially unbalanced situation. Both stocks are influenced by inflow and outflow. Participants are able to influence the inflow by deciding on the hiring percentage for women, which immediately influences the hiring percentage of men. The outflow of the stock depends upon the value of the stock at a particular moment in time. When the number of female full professors increases, the outflow of female professors increases proportionally, and when the number of male full professors decreases, the outflow of male professors also decreases proportionally. Participants thus need to realize that, in this case, the stock of female full professors follows a growth path that slows down over time.

Sample

The experiment was conducted with two groups of students at a Dutch university. All students followed a bachelor’s course on personnel policies as part of a teaching program on Human Resource Studies. Because students were following a higher education curriculum and had already finished introductory courses on Human Resource (HR) studies, we may consider them basically skilled and interested in addressing dynamic problems in the field of personnel policies. They did not receive an incentive for answering correctly, but received feedback on the performance as part of their course program.

The participants answered the open questions of the (English) task in Dutch or English. Not all participants answered the questions completely. Information on year of birth, study, gender or recommended hiring percentage was

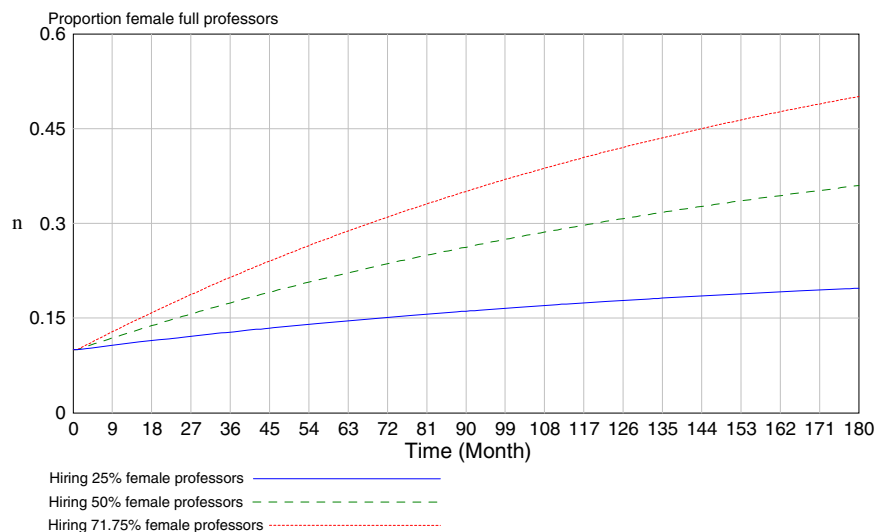
missing for 24 participants. We decided to exclude the participants who had information missing on at least one of the variables mentioned. As a result, data from 133 of the original 157 participants remained for further analyses. The mean age of the group of 133 participants included in the analysis was 20.8 years, and 76.7 percent of them were women. The participants were studying in the undergraduate bachelor program in HR Studies (43.6 percent), the undergraduate premaster program in HR Studies (24.1 percent) or the bachelor program in Organization Studies (32.3 percent).

Results

This section presents the estimated hiring percentages of the participants and the way they substantiated their decision. We simulated the task in Vensim as visualized in Fig. 2 and determined the hiring percentage for women needed to reach the 50:50 target in 15 years. According to this simulation, a hiring percentage of 71.75 percent female full professors per year is needed to reach an equal balance between male and female full professors in 15 years.

First, we describe two variables: the hiring percentage for women as estimated by the participants (further denoted as the estimated percentage); and the difference score between the estimated percentage and the correct percentage (further denoted as the difference score), calculated by subtracting 71.75 from each estimated percentage. A negative difference is an underestimation, whereas a positive difference is an overestimation. Kolmogorov–Smirnov tests showed for both variables that the data were not normally

Fig. 2. Simulation results of different hiring proportions of women



distributed: estimated percentage, $Z=0.13$, $p<0.001$; difference score, $Z=0.13$, $p<0.001$.

Statistics on the central tendency and dispersion are presented in Table 2. The median for the estimated percentage is 50 percent. The estimates range from 2.0 to 100 percent, and half of all participants recommended a percentage between 29 (Q1) and 70 (Q3) percent. The median for the difference score is -21.8 percent. The scores range from -69.8 to 28.3 percent, and half of all participants recommended a difference score between -42.8 (Q1) and -1.8 (Q3) percent. The results of a sign test show that the median for the difference score significantly deviates from zero: the median is lower than the correct percentage by 21.8 percent.

In summary, a large majority of the participants underestimated the hiring percentage needed to accomplish the set target (see Fig. 3). If we take a margin of 2 percent around the correct percentage to reach the 50:50 target in 15 years, 74.4% of the participants estimated a lower percentage, and 19.5 percent of the participants estimated a higher percentage than needed. Only 6.1 percent of the participants fell within the margin of the correct percentage. A Kruskal–Wallis test showed that the percentage estimated significantly differed ($H(2) = 76.844$, $p < 0.001$) between these three groups of participants. The median percentages were 47.00 for the participants who estimated a lower percentage than needed, 70.50 for the participants who estimated an almost equal percentage and 82.48 for those who estimated a higher percentage than needed.² Moreover, 38.3 percent of the participants estimated a hiring percentage below 50 percent, by which the target would never be met at all.

Analysis of arguments

To reveal patterns in the way participants substantiated their decisions, we performed a qualitative content analysis of 100 substantiations. (The other 33 participants estimated a percentage without substantiating it.) The written material was transcribed and coded by two researchers. We used inductive coding (Strauss and Corbin, 2008) to ensure the inclusion of themes that we did not expect beforehand. In Table 3, we present the types of substantiations we distinguished. In deliberation between the researchers during a series of coding rounds, we reached a consensus on three main types of substantiation among a total of 86 substantiations, i.e. calculations (with three subcategories), intuitive estimations and ethical argumentations (Table 3). We were unable to code the remaining 15 substantiations because they were incomplete or incomprehensible, and we excluded them from the qualitative

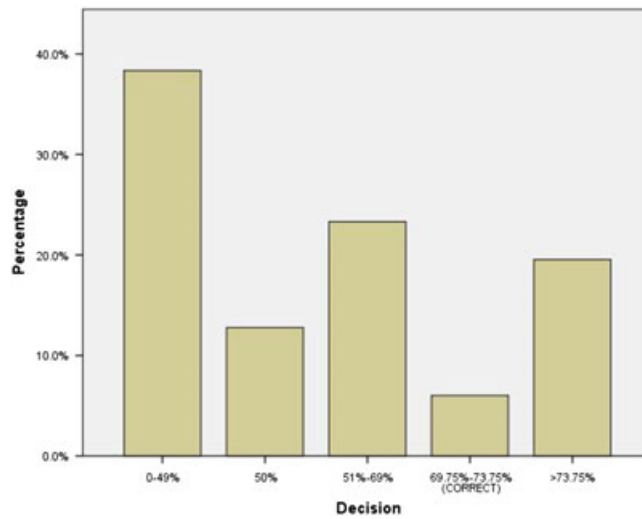
²This result should be cautiously considered because only eight participants estimated a percentage that is almost equal to the correct percentage needed to reach the target. Excluding these eight participants, an additional Mann–Whitney U -test showed that the estimated percentage of those who estimated a percentage lower than needed (Mdn = 47.04) was significantly lower than the estimated percentage of those who estimated a higher percentage than needed (Mdn = 82.84), $U = 2547.00$, $p < 0.001$.

Table 2. Estimated hiring percentage and the difference score ($N = 133$)

	Estimated percentage	Difference score
Median	50%	-21.8**
Min.; max.	2%; 100%	-69.8; 28.3
Q1	29%	-42.8
Q3	70%	-1.8

** $p < 0.01$ (sign test, two-sided).

Fig. 3. Histogram of the estimated hiring percentage



dataset. Moreover, we decided not to further involve six substantiations related to ethical concerns in our analysis. Although the task stipulated the use of a gender quota and although participants were supposed to agree with the goal of reaching a 50:50 gender balance, the participants who used this type of substantiation argued that they (dis)agreed with the use of gender targets or quota.³ We considered these arguments to be irrelevant for exploring how people understand the dynamic system underlying the task. As a result, the subsequent analysis is based on the remaining 79 substantiations.

Calculations

The first type of substantiation was based on a calculation (defined as a range of numbers with a plus, minus, times and division sign). We also included

³For example, participants with ethical argumentations argued that “[t]hey should choose the best qualified men and women for the position” (this respondent estimated a hiring percentage of female full professors of 50 percent). The six participants using this substantiation did not perform well on the task. They chose hiring percentages between 5 percent and 50 percent, which is not sufficient to reach the target in time or even at all.

Table 3. Substantiations
($N=85$)

Types of substantiation	N (%)	Correct answer (%)
Calculations:	41 (48)	
1. Correlation heuristic	7 (8)	
2. Filling-the-gap heuristic	24 (28)	Absent
3. Pipeline heuristic	10 (12)	
Intuitive estimations:		
Compensation is needed	38 (45)	6 (7)
Ethical argumentations:		
Gender targets should not be applied	6 (7)	Absent
Total	85	6 (7)

substantiations with similar content in which numbers were verbally added, subtracted, duplicated or divided.

Category 1: stock behaves like the flow

The first category of calculations substantiated the estimated percentage (flow) by referring to the increase in the stock needed to achieve the target. These participants used the following type of reasoning: the percentage is now 10 percent; it needs to go up to 50 percent; an increase of 2.6 percent per year is thus required; and the target will be achieved by a hiring percentage equal to the percentage the stock needs to increase per year. An example follows:

40% extra women in 15 years. $40/15 = 2.6\%$ extra women/year [The respondent estimated a hiring percentage of 2.6%]

All participants (8 percent) using this substantiation recommended a hiring percentage below 10 percent. In reality, a hiring percentage below 10 percent would lead to a decrease rather than an increase in the present proportion (10 percent) of female full professors. In that sense, these participants performed the worst of all. By assuming the system output (proportion) behaved like the system input (hiring percentage), these substantiations reflected previously identified patterns of stock–flow failure framed as the “correlation heuristic” (Cronin *et al.*, 2009; Sterman, 2010).

Category 2: inflow fills the gap between the stocks

The second category of calculations relates to the gap between the stocks of female and male full professors. These participants divided this gap over the number of expected vacancies during the period to estimate the hiring percentage. We call this reasoning pattern the “filling-the-gap heuristic”, i.e. a reasoning pattern in which inflow was erroneously assumed to affect the

gap between two stocks rather than the stocks themselves. From the 24 participants who used this type of substantiation, three categories of calculations can be distinguished: those that did not take any outflow into consideration (2a), those that underestimated the outflow of female full professors (2b) and those that overestimated the outflow of female full professors (2c).

Without outflow

Eight participants (9 percent) assumed that the gap between the two stocks of 400 could be fixed by hiring 400 female full professors divided by the turnover of 1050 full professors over 15 years. Male full professors were assumed to flow out, whereas the outflow of female full professors was not taken into consideration.

Male professors: 900, female: 100. $\Rightarrow 900 - 400 = 500$; $100 + 400 = 500$. $400:15 = 26$ female professors should be hired annually. $26:70 = 37\%$. [The respondent estimated a hiring percentage of 37%.]

The participants who used this substantiation advised a hiring percentage between 2.7 percent (three participants) and 48 percent (one participant), with most estimating approximately 38 percent (four participants). The hiring percentages in this category seem less divergent than they appear at first sight because the participants referred to 2.7 percent of the 7 percent turnover, which could be converted to 38.6 percent. In all cases, the estimated percentage is too low to reach the target. In this sense, these participants performed poorly, but not as poorly as the participants who used the argument that the stock behaves like the flow.

With outflow underestimation

Ten participants (12 percent) assumed that the discrepancy between the two stocks of 400 could be fixed by hiring 400 female full professors divided by the turnover of 1050 full professors, complemented by the initial turnover percentage of female full professors. An example follows:

$\text{equal} = 500/500$. From 100 to 500 in 15 years $= 400/15 = 26.67$ women each year. Turnover is 70 of which 7 are women (10% of 70). $26.67 + 7 = 33.67$ women needed each year. 70 vacancies. $33.67 = 48.09\%$ of the 70 vacancies each year. [The participant estimated a hiring percentage of 48%.]

The participants who used this substantiation recommended a hiring percentage between 5 percent and 69 percent, which is too low to reach the target in 15 years. These percentages differ because the outflow that should be compensated for is sometimes calculated on the basis of the turnover, sometimes calculated on the basis of the stock or just estimated low. By fixing the outflow

of female full professors at the level of the starting situation, these participants failed to take into account the fact that the outflow of female full professors increases rapidly when the two stocks become more balanced. We dubbed this reasoning pattern the “fixed outflow heuristic”, i.e. a reasoning pattern in which the outflow of a stock is erroneously assumed to be fixed at a certain level. Used within the “filling-the-gap” heuristic, this heuristic leads to an underestimation of the inflow needed to balance the two stocks. Nevertheless, the underestimation was limited for these participants compared with other groups of participants. One participant in this group estimated the correct hiring percentage.

With outflow overestimation

Six participants (7 percent) assumed that the difference between the two stocks of 400 could be fixed by hiring 400 female full professors divided by the 1050 turnover in 15 years, complemented by an assumed 50% leaving percentage of female full professors. An example follows:

Per year, 26.67 more female professors need to be hired (400:15 years). To keep it 50/50 per year, 35 men and 35 women need to be hired. $35 + 26.67 = 61.67$. $61.67/70 = 0.8809 = (\text{about}) 88\%$. [The respondent estimated a hiring percentage of 88%.]

The percentage of female full professors leaving in the first period would be much lower because it equals the proportional value of the stock of female full professors (initially 10 percent). Within the “filling-the-gap heuristic”, fixing the outflow of the two stocks at equal levels leads to an overestimation of the inflow needed. The hiring percentage of 88 percent (with one exception of 6.16 percent) derived from this substantiation is higher than that needed to reach the target in time. The hiring percentage of 6.16 percent is less divergent than it appears at first sight because this participant referred to 6.16 percent of the 7 percent turnover, which can be converted to 88 percent. This was the only group of participants who seriously overestimated the hiring percentage needed.

Category 3: stock behaves like a discrete pipeline

The third category of calculations assumes the stock to behave like a discrete pipeline. These participants assumed that all full professors will leave exactly 15 years after being hired. As a result, these participants argued that an annual allocation of half of the vacant positions for women would result in the realization of the desired situation in the same period because all available positions would be taken by new professors in 15 years. An example follows:

When there is 7% professor vacancy every year, it will be 1050 professors in total. So, if you hire 50% female professors that will result in a 50–50 gender balance [in 15 years—AUTH]. [The participant estimated a hiring percentage of 50%.]

The 10 participants who used this substantiation (11%) all advised a hiring percentage of 50 percent. However, according to the task, the turnover of full professors is *on average* 7 percent per year. This means that, for example, some full professors leave after 2 years, and others leave only after 20 years. Therefore, the task assumes a continuous rather than a discrete delay, which has consequences for the dynamic system (cf. Größler and Zock, 2010). If a hiring percentage of 50 percent female full professors is adopted and if the turnover is on average 7 percent per year, the percentage of female full professors will increase rapidly in the beginning, but the growth will slow down over time because more female professors would also be leaving. We call this misunderstanding the pipeline heuristic, i.e. a reasoning pattern in which a discrete delay in reaching a target is erroneously assumed. The participants who used this heuristic seriously underestimated the hiring percentage.

Intuitive estimations

The second type of substantiation refers to a more intuitive estimation that the unbalanced character of the stocks should be compensated for by a larger inflow of women relative to men until the target is reached. We label these substantiations intuitive estimations because the large majority of them refer to the need for compensation in general terms without underpinning it with calculations.⁴ They argue that, for a certain period, more women need to be hired than men. Examples follow:

Hire a bit more women than men every year because more men [than women] are leaving. It will slowly become more equal without showing too much that you want this number to be equal. [The participant estimated a hiring percentage for women of 60%.]

80%, because you will reach the 50-50 division quickly this way and you still give men an opportunity in the selection procedure. [The participant estimated a hiring percentage for women of 80%.]

The 38 participants who used these substantiations (45 percent) recommended percentages above 50 percent (with the exception of two, who recommended 4 percent and 6.6 percent). The exceptions are less divergent than

⁴Nevertheless, five participants underpinned references to the need for compensation with calculations, and two participants entirely supported this type of substantiation with a calculation. However, their reasoning patterns best fitted within this group.

one should assume because the participants referred to 4 percent and 6.6 percent of the 7 percent turnover, which can be converted into 57 percent and 90 percent. The participants who used intuitive estimations performed best on this decision task, as they were the only ones to give the correct estimations of the hiring percentage to solve the task (with a 2 percent margin).⁵ The ones with the correct hiring percentage used similar arguments to the other participants with intuitive estimations, sometimes roughly estimating how much above 50 percent the hiring percentage should be:

You need to hire more women. [It is] also important [to take into account] that women may leave, so 50/60% may be too low, I guess. [The participant estimated a hiring percentage of 70%]

70 open positions [should be] filled within 15 years by hiring 20% more women than men. [The participant estimated a hiring percentage for women of 70%.]

This finding is contrary to earlier research on dynamic decision making, in which people are considered to have a poor intuitive understanding of dynamic systems and in which intuitive understanding is considered to hinder individuals' ability to correctly estimate dynamic behavior (Sterman and Booth Sweeney, 2002). With stock–flow tasks of modest dynamic complexity like the faculty gender balancing task, we can hardly expect participants to come up with the correct answer of 71.75 percent (with a 2 percent margin roughly from 70 to 74) without computer simulation. Intuitive understanding that the hiring percentage should be considerably over 50 percent for a given period, however, shows a basic understanding of the dynamic system and is preferable to detailed calculations that only take isolated parts of the dynamic system into account.

Discussion and conclusions

Our analysis offers insight into the reasoning patterns on which people rely to perform decision tasks with modest dynamic complexity. These reasoning patterns reveal the difficulties people have in understanding the dynamic system underlying balancing two stocks in an initially unbalanced situation. Analysis of the underpinning of the estimations indicates that individuals use intuitive estimations (nearly half of the participants) or calculations based on heuristics (half of the participants), i.e. the correlation heuristic, the filling-the-gap heuristic and the pipeline heuristic. The correlation heuristic, which was identified earlier to explain low performance in dynamically simple tasks, such as the department store task (Sterman, 2010, p. 328; Pala and

⁵The other two participants that fell within the margin of the correct answer did not substantiate it at all.

Vennix, 2005), was used by a small percentage of the participants who underpinned their decision (8 percent). Other heuristics were used more often. We identify a filling-the-gap heuristic (used by 28 percent of the participants who underpinned their decision), where the participants erroneously assumed that the flow affects the gap between two stocks rather than the stocks themselves. Finally, we identify a pipeline heuristic (used by 11 percent of the participants who underpinned their decision), where the participants assumed that the delay in reaching the target was discrete rather than continuous. This finding indicates a mental model involving a “first in, first out” (FIFO) approach. The use of such an approach is understandable because it requires much less mental effort to estimate the effects of a discrete delay than a continuous delay. By identifying how (a combination of) these heuristics relates to a failure in understanding the dynamic system underlying the task, we reveal the variety in individuals’ erroneous reasoning patterns.

A relevant finding for the system dynamics community is that the participants who supported their decision with intuitive estimations (45 percent of the participants who underpinned their decisions) performed best on the task in the sense that the correct estimations were only given by this group. These participants guessed that the inflow temporarily needs to be higher than 50 percent to compensate for the imbalance between the stocks. This indicates that individuals may perform better in decision making in tasks of modest dynamic complexity, such as that simulated in the faculty gender balancing task, if they develop an intuitive understanding of the problem than if they make detailed calculations that only take isolated parts of the dynamic system into account. Nevertheless, an intuitive understanding seems a necessary but not a sufficient condition to solve the task correctly; only six out of 38 participants with intuitive estimations correctly estimated the hiring percentage. The others showed a more general understanding of the dynamic system by arguing that the correct answer should be above 50 percent.

Sterman and Booth Sweeney argued that individuals’ intuitive understanding of even the simplest dynamic problem is poor (2002, p. 234). Our results suggest that individuals’ understanding of dynamic systems with modest dynamic complexity is also poor but that the substantial minority of individuals who are able to correctly estimate the dynamic behaviors use intuition rather than heuristics to support their understanding. We wonder whether this intuition results from experience with dynamic systems with behavioral similarity that has been unconsciously processed. We suggest that supporting intuitive understanding of complex dynamic problems, for example by comparing between dynamic systems in different fields, may be the way to improve decision making on these issues.

This exploratory study has consequences for research and practice. Our findings further develop Sterman and Booth Sweeney’s (2002) statement regarding the difficulties people have in understanding the inertia of dynamic systems in general by indicating which heuristics cause a failure in

individuals' ability to understand what is needed to balance two unbalanced stocks. A meaningful follow-up would be a survey study testing the presence of these heuristics in other, dynamically more complex decision tasks. Fischer and Degen (2012) suggest that a verbal representation of stock–flow tasks may increase the understanding of stock–flow problems. With the faculty gender balancing task, we presented information verbally rather than graphically to support the participants' understanding of the dynamic problem. Our results indicate that a majority of participants failed to understand the dynamic system underlying balancing two unbalanced stocks, despite the use of a verbal representation. Gonzalez and Wong (2012) suggest that comparing stock–flow problems with other problems with behavioral similarity may increase the understanding of stock–flow problems. Considering the relative success of the estimations based on an intuitive understanding, we recommend that follow-up research use behavioral similarity to improve the intuitive understanding of the dynamic systems underlying stock–flow tasks.

The study also has practical consequences. By using the real-world problem of setting gender targets in personnel policies, we showed how difficulties in understanding the underlying dynamic system potentially affect decision making on the issue. Cognitive reasoning with the correlation heuristic or the filling-the-gap heuristic combined with a fixed outflow heuristic may induce individuals to decide on an inflow that is too low ever to reach a balance between the stocks. Decision makers need to take into consideration that the increasing outflow of the underrepresented group slows down the speed by which two stocks in an unbalanced situation can be balanced. A limitation regarding the practical consequences of our study is that the task was performed by students of HRM rather than by professionals working in the field. An experiment with a simpler stock–flow task, the department store task, showed that participants with an average of 5 years' practical experience did not perform significantly better than participants without practical experience (cf. Sterman, 2010). However, the influence of practical experience still needs to be tested for the faculty gender balancing task. Because our experimental group was being educated for positions in HRM, our results suggest this group of students needs to be made aware of how stock–flow failure could negatively affect their decision making in personnel policies. We used the results of the experiment to show the students the potential failures in decision making regarding this issue.

To conclude, this study aimed to contribute to the understanding of decision making on dynamically complex problems by exploring how people understand the dynamic system underlying balancing two stocks with negative feedback loops. An experiment with the faculty gender balancing task showed that participants had difficulty correctly estimating the inflow needed to balance two unbalanced stocks. We found that a large majority of the participants underestimated the hiring percentage needed to balance the two stocks in time: three-quarters of the participants recommended a hiring percentage that was lower than that needed to reach the target in the target year. Moreover,

two-fifths of the participants even estimated a hiring percentage by which the stocks would never be balanced at all. Participants who supported their decisions with calculations used three types of heuristics, reasoning patterns on which they erroneously relied to assess the dynamic behavior. We found that the participants who used intuitive estimations to support their decisions performed relatively better on the task. This indicates that fostering an intuitive understanding of dynamically complex problems may be the way forward to improve decision making on these issues.

Biographies

Inge Bleijenbergh holds a PhD from the Vrije Universiteit Amsterdam in Social and Cultural Sciences. She is an Associate Professor of Research Methods at Radboud University Nijmegen, The Netherlands, and teaches in the European Master Programme in System Dynamics. Her main research areas are related to group model building and gender and diversity in organizations.

Jac Vennix holds a PhD from Radboud University Nijmegen in Social Sciences. He is Professor of Research Methods at the Institute for Management Research at the Radboud University Nijmegen, The Netherlands. He is one of the founding fathers of the European Master Programme in System Dynamics. His main research areas are related to group model building and decision making on complex problems.

Eric Jacobs holds a PhD from the Radboud University Nijmegen in Social Sciences. He is an Assistant Professor of Research Methods at the Institute for Management Research at the Radboud University Nijmegen, The Netherlands. His main areas of research are related to decision making and information exchange.

Marloes van Engen holds a PhD in Social and Behavioral Studies from Tilburg University. She is an Assistant Professor of Human Resource Studies at Tilburg University. Her main areas of research are related to gender and diversity in organizations and leadership.

References

- Bleijenbergh I, Vennix J, Jacobs E, van Engen M. 2011. Reducing the gender gap: biases in understanding delays in personnel policies. In *Proceedings of the 2011 International Conference of the System Dynamics Society*, Cambridge, MA.
- Booth Sweeney L, Sterman JD. 2000. Bathtub dynamics: initial results of a systems thinking inventory. *System Dynamics Review* **16**(4): 249–286.
- Charter Talent to the Top. 2016. *Ondertekenaars* [Signers]. Available: http://www.talentnaardetop.nl/Home_EN/Charter/Signatories [25 May 2016].

-
- Cronin MA, Gonzalez C, Sterman JD. 2009. Why don't well-educated adults understand accumulation? A challenge to researchers, educators and citizens. *Organizational Behavior and Human Decision Processes* **108**(1): 116–130.
- Dudley R. 2007. The equity supply chain: is it the cause of so few women in management and leadership positions? In *Proceedings of the 27th International Conference of the System Dynamics Society*, Boston, MA.
- Fischer H, Degen C. 2012. Stock–flow failure can be explained by the task design. In *Proceedings of the 30th International System Dynamics Conference*, St Gallen, Switzerland.
- Gerritsen M, Verdonk T, Visser A. 2012. *Monitor Vrouwelijke Hoogleraren 2012*. Available: http://www.stichtingdebeauvoir.nl/wp-content/uploads/Monitor_Vrouwelijke_Hoogleraren_2012.pdf [6 May 2014].
- Gonzalez C, Wong H. 2012. Understanding stocks and lows through analogy. *System Dynamics Review* **28**(1): 3–27.
- 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. In *Proceedings of the 20th International Conference of the System Dynamics Society*, Palermo, Italy.
- Kampmann C, Sterman JD. 2014. Do markets mitigate misperceptions of feedback? *System Dynamics Review* **30**(3): 123–160.
- Korzilius H, Raaijmakers S, Rouwette E, Vennix J. 2014. Thinking aloud while solving a stock–flow task: surfacing the correlation heuristic and other reasoning patterns. *Systems Research and Behavioral Science* **31**: 268–279.
- Merens A, Hartgers M, van den Brakel M. 2012. *Emancipatiemonitor 2012*. 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. In *Proceedings of the 20th International Conference of the System Dynamics Society*, Palermo, Italy.
- Pala Ö, Vennix JAM. 2005. Effect of system dynamics education on systems thinking inventory task performance. *System Dynamics Review* **21**(2): 147–172.
- Phuah TLK. 2010. Can people learn behaviours of stocks and flow using their ability to calculate running total? An experimental study. In *Proceedings of the 28th International System Dynamics conference*. Seoul, Korea.
- Sterman JD. 2002. All models are wrong: reflections on becoming a systems scientist. *System Dynamics Review* **18**(4): 501–531.
- Sterman JD. 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.
- Sterman JD, Booth Sweeney L. 2007. Understanding public complacency about climate change: adults' mental models of climate change violate conservation of matter. *Climatic Change* **80**: 213–238.
- Strauss A, Corbin J. 2008. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. Sage: Thousand Oaks, CA.
- 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.

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

Appendix: The faculty gender balancing task[†]

In this task, you take on the role of the chair of a university board. It is the board's objective to reach a gender balance in the number of full professors at that university in fifteen years.

The situation at the university

Currently, the university has 1,000 full professor positions, of which 90% are occupied by men (i.e., 900 male professors) and 10% by women (i.e., 100 female professors). In contrast, the proportion of female and male students has been 50-50 for the past few decades. The board of the university has decided to drastically increase the number of female full professors. Thus, the task for you as the chair is to realize a 50-50 distribution of women and men in full professor positions at the university in fifteen years.

Considering the financial situation of the university, the total number of full professor positions cannot be increased, and it will remain at 1,000 for the years to come. The change therefore has to occur within the available number of full professor positions, meaning that only after a full professor leaves (i.e., there is a vacancy) can a new professor be appointed to that position.

The turnover of full professor positions is, on average, 7% per year, resulting in approximately 70 vacancies annually, which can be filled by new professors. This 7% holds for both male and female full professors. This average turnover of full professors has been the same since the university was first established. You may assume that full professors' departures are equally distributed over the different age groups.

The strategy

To achieve the goal, the board decided to implement a quota. As the chair, you need to tell the directors of schools what percentage of women they should hire annually for these vacant positions.

[†] Thanks to Erling Moxnes and editors of SDR for their comments on earlier versions of the task.

Additional information

First, note that there are enough qualified female and male candidates for the positions available, whatever percentage you choose. Second, remember that as the chair, you fully agree with the goal of reaching a 50-50 gender balance and are responsible for implementing this goal. Third, to prevent fluctuations in hiring policies, you can decide on only one percentage, which is used annually to distribute male and female professors over the vacant positions until the 50-50 gender balance is reached.

Decision

Now, please think carefully about what percentage of female full professors you would like the schools to hire to reach a gender balance in fifteen years.

My decision on an annual hiring percentage of female full professors: ___%
(The rest will be male full professors.)

Please briefly explain below why the objective of equal gender balance will be achieved in fifteen years, given the percentage you have selected.

.....

Your year of birth: 19__

Your area of study:

Woman/man

Thanks for your participation!