The Effect of the Time Unit on Software Development Effort Estimates

[No author names, use paper id if you have one]

***Abstract*—Estimates of software development effort are frequently inaccurate and over-optimistic. In this paper we describe how changes in the granularity of the unit of estimation, e.g., work-days instead of work-hours, affects the effort estimates. We describe four psychological mechanisms, how they interact and discuss the expected total effect of higher granularity units on effort estimates. We argue that the mechanisms in general imply that higher granularity effort units will result in higher effort estimates, e.g., that estimating software development work in work-days or weeks will lead to higher estimates than when estimating in work-hours. A possible implication of this predicted effect is that, in contexts where there is a tendency towards under-estimation, estimation in work-days or weeks instead of work-hours leads to more realistic estimates.**

Keywords—cost estimation, human judgment, psychology

I. Introduction

Empirical evidence documents a tendency toward cost and effort overruns in software projects. On average, this overrun seems to be around 30 percent [1] and one out of six software projects seems to cost more than twice the initial estimate [2]. Furthermore, comparing the estimation accuracy of the 1980s with that reported in more recent surveys suggests that the estimation accuracy hasn't changed much since then. Estimation methods haven't changed much either. In spite of an extensive research on formal estimation models, the dominating estimation method is still expert estimation of effort required for activities listed in a work break down (WBS) structure. We summarize state-of-practice and research-based knowledge of effort estimation in [3, 4].

While there has been much research on formal effort estimation models, there has not been much research on expert judgment-based effort estimation. One topic, part of expert judgment, not studied at all, is how the unit used for estimating the required amount of work-effort affects the software professionals' estimates. We have through our software project research and practice as software professional experienced that software professionals tend to estimate software project activities in work-hours, but that work-days and work-weeks are also sometimes used. Could it be the case that estimating in work-hours instead of for example work-days affects the effort estimates and, perhaps, even contributes to the observed tendency towards too low effort estimates?

This paper describes four mechanisms, which all have in common that they predict that the chosen unit will have implications for the effort estimates. We explain and apply these mechanisms to effort estimation contexts with the goal of better understanding and improving judgment-based effort estimation work.

II. How the Unit may Affect Estimates

A. Numerosity

In 1928 Fisher Irving wrote a book called "The money illusion" [5]. The book points out that people's judgments are affected by both the nominal and the real values of numbers. Assume that we receive 5% income raise and the inflation is 6%. We then have a net loss of 1% in income. Compare this with the situation where we have no income raise, i.e., 0% income raise, and the inflation is 1%. We have as before a net loss of 1% in income, but most people will react much more negatively in the second situation. Although we know that what really matter is what can buy for the income, we tend to get affected by the nominal value of the income. Higher numbers (higher numerosity) typically give us the feeling that there is more of a quantity than the same quantity when using a unit that gives lower numbers.

This numerosity-based error in human judgment has been documented in several domains. One feels for example that goods become less expensive when buying in a currency that has lower numerical values. As expected from the numerosity effect there was an increase in donations given by church visitors when Italia went from Lira, with high nominal values, to Euros, with much lower nominal values [6].

For time units there is evidence that 365 days may feel longer than 12 months, which in turn may feel longer than 1 year. The studies reported in [7], for example, found that the likelihood of starting a diet was higher when framed as a one year plan rather than a 365-days plan. Interestingly, the effect of the unit describing the length of the plan increased substantially with increased personal relevance of dieting. This suggests that there are a number of moderating factors determining the strength of the numerosity effect.

If people feel that 365 is longer than one year, due to the numerosity effect, they may also expect that it is possible to complete more work in 365 days than in one year. This, in turn, may imply that estimating the effort of a software development task in a low granularity work unit, such as work-hours, should lead to lower estimates than when estimating the same task in higher granularity work units, such as work-days.

B. Unitosity

Not only the differences in numerosity caused by different units, but also the unit itself may affect the estimates [8, 9]. This may be due to conversational norms [10], e.g., the norm that we usually request the estimates of smaller amounts using lower granularity and higher amounts using higher granularity units. We do, for example, typically not request the effort of a small task in man-years, but rather in work-hours or minutes. The unit of the estimation request may therefore be interpreted as an indicator of the expectation of the person who asks for the estimate. This expectation has been documented to have an effect on the estimation work, even when we know that the expectation is not based on high competence [11].

While never been studied in relation to effort estimation unit, we have previously documented that other types of request formats indicating expectations affect the effort estimates. We found, for example, that the request to estimate “a minor modification” resulted in much lower estimate than the request to estimate “a new functionality” for exactly the same software development specification [12]. Similarly, a question, revealing an expectation of unrealistically low effort usage, from a client with no technical competence, resulted in much lower effort estimates, in spite of explicitly being told to disregard the request as valid information [13].

Our expectation is consequently that when receiving a request to estimate effort applying a low granularity unit, such as work-hours, we will be affected to think that the task is smaller compared to the situation when requested to estimate applying a higher granularity unit, such as work-days.

C. Construal Level

The construal level theory aims at explaining the relation between “psychological distance” to the subject of our thinking and the abstraction level of the thinking. Psychological distance may, amongst others, increase with distance in time (next year vs. today activities), physical distance (objects far away vs. here), social distance (people in another organization vs. in my organization) and hypothetical distance (something that will happen vs. something that may happen). The construal level theory predicts that higher psychological distance leads to higher construal level, i.e., higher abstraction level and more goal-oriented thinking [14]. The relevance of distance in time for abstraction and goal-orientedness of thinking has been documented time in for example [15, 16]. The relation between construal level and the effect of numerosity and unitosity is reported in [9]. In that paper the authors suggest that when manipulating (priming) the participants to higher construal level thinking, they focused more the unit and less on the number, i.e., the unitosity mechanism had larger effect than numerosity mechanism.

There are diverging results on the effect of construal level on effort or time estimates. The experiments reported in [17, 18] find that an increase in construal level increased the performance estimates. The experiments in [19], on the other hand, report that the performance estimates increased with increase in construal level for simple tasks only. The estimates decreased, however, for more complex tasks.

There are, as far as we know, only one study on the effect of choice of estimation unit on the construal level. This study report that scales with higher granularity increased the psychological distance, i.e., that “… large scale leads to big picture thinking” [16].

Based the previous results we hypothesize that the use of a higher granularity effort unit will lead to higher construal level thinking. The results in [19] implies that higher construal level thinking leads to higher estimates for easy and lower estimates for more complex tasks. This means, for example, that estimates in work-days should give higher estimates for simple software development tasks and lower estimates for complex projects than estimates in work-hours.

D. Fluency

We may process mental work, such as estimating required work-hours of software development activities, with ease or we may find it difficult and inefficient. The subjective experience of ease with which we complete mental processes is what we mean by cognitive fluency [20].

Our judgments may be affected by the fluency of the mental process producing the judgment. We may for example use the fluency of the mental process trying to understand a requirement specification as an indicator of how difficult the software work will be. In [21], for example, it is reported that easy-to-read instructions, high font readability, produced lower time estimates than hard-to-read instructions, low font readability, for the same task. In [8] the default (most fluent) unit for judging product characteristics led to the most favourable evaluations and explain this as caused by processing fluency.

Applied on or examination of the effect of different estimation units, we hypothesize that if an effort unit affects the fluency of the estimation work, e.g., by inducing a unit that is unnatural or not corresponding with the unit of the experience, this will increase the effort estimate. A decrease in fluency may, for example, be the case when estimating a small task using high granularity time units.

III. The Total Effect

The discussed mechanisms do not give the same predictions of what will happen with the effort estimates when using a high rather than low granularity effort unit, e.g., work-days instead of work-hours. The total effect of choice of unit on effort estimates will depend on the mechanisms relative effect sizes, which in turn may depend on contextual factors and moderating variables, and possible also non-included mechanisms of relevance. Possibly, the total effect will only be possible to know when empirically studying it in relevant contexts.

In spite of the above challenges of knowing the total effect, it may be useful to integrate the four described mechanisms and what we know about moderating variables in a, admittedly, very simplified picture This is what we have tried to do in the integrated model in Figure 1, where a “+” means that we would expect a positive and a “-“ a negative correlation between values.

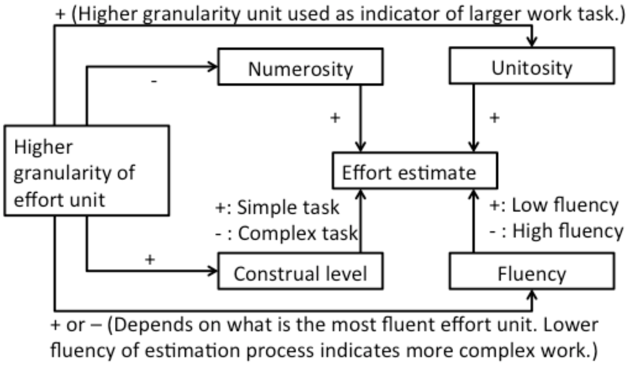

Fig. 1. The integrated model of the mechanisms

Figure 1 suggests that the use of a higher granularity effort unit is expected to lead to higher estimates in many, perhaps most, contexts. It is only if the task is complex and/or the higher granularity effort unit leads to lower estimation process fluency that higher granularity unit may lead to lower effort estimates. In addition, for that to happen the effect sizes of construal level and fluency have to be larger than those of numerosity and unitosity. Our experience is that construal level and fluency effect sizes in effort estimation work are not very strong (we have several failed attempts to find any statistically significant effect) and we therefore expect that the general effect of a higher granularity effort unit would be higher effort estimates.

IV. EMPIRICAL DATA

A. Experiment 1

As a first pilot of the effect of the time unit on the effort estimates, we asked 10 participants on an software development effort estimation lecture to estimate the number of second (Group 1) or minutes (Group 2) they would need to solve a small, well-defined task, i.e., a paper folding (origami) task. We had previously tested this origami task on 16 other students and found that they on average (mean and median about the same) needed 280 seconds. The numerosity and unitosity effects would predict higher estimates in minutes than in seconds. The effect of the construal level and fluency theory was not clear, but we expected the task to be considered as simple and that both seconds and minutes would be natural units for estimating this kind of task, i.e., no large effect from difference in fluency. In total, we therefore hypothesized (H1) that:

H1: *The estimates in seconds will on average be lower than those in minutes.*

The collected estimates supported H1. The mean effort of those estimating in second was as high as 924 seconds (median of 600 seconds), while the mean effort of those estimating in minutes was 142 seconds, see Fig. 1. A one-sided t-test of difference in mean values gives p=0.02.

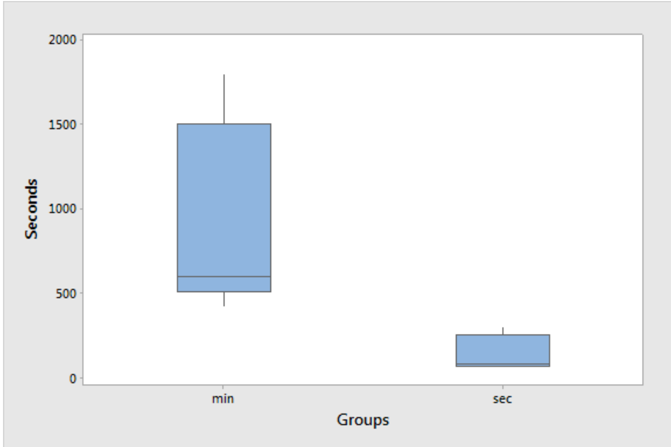

Fig. 2. Comparison of estimates in minutes and seconds

With only few participants we should be very careful about claims about who were the most accurate in their estimates. Comparing with the actual time of people conducting the task suggests that most of those estimating in minutes were quite over-pessimistic. Most of those estimating in seconds were over-optimistic, but likely to be more accurate than those estimating in minutes.

B. Experiment 2

Our second examination of effect of the unit of the effort estimates was conducted with twenty-two experienced project managers in a large software development company. They were randomly divided into two groups. In one group the participants should estimate a software development project in work-hours and in the other group in workdays. The participants got only 10-15 minutes on the estimation task and were asked to base the estimates on similarities with previously completed projects, i.e., so-called analogy-based estimation. The participants were also asked to describe their estimation process. Those who estimated in work-days were told to give a conversion factor between work-days and work-hours. The short time for estimation was expected to strengthen the effect of the experimental manipulation, see for example [22].

The numerosity and unitosity mechanisms are, as before, expected to make estimates in work-days higher than estiamtes in work-hours. We did not expect strong effects from construal level or fluency, as this was a medium complex task and both work-hours and work-days were common estimation units. We therefore hypothesized (H2) that:

H2: *The estimates in work-hours will on average be lower than those in work-days.*

The collected estimates supported our hypothesis. The men effort estimate of those in the work-hours group was 84.5 work-hours (median of 77 work-hours), while the corresponding mean estimate was 266 work-hours (median 115 work-hours or 17.5 work-days) for those in the work-days group.

A one-sided t-test test of difference in mean values gave p=0.06. Fig. 2 displays a boxplot with the estimates for the two groups.

In the analysis we removed two effort estimates. These two participants were instructed to estimate in work-hours, but described that they had estimated in higher granularity units. Including these two observations, categorized as using work-days, would have further increased the difference in estimates.

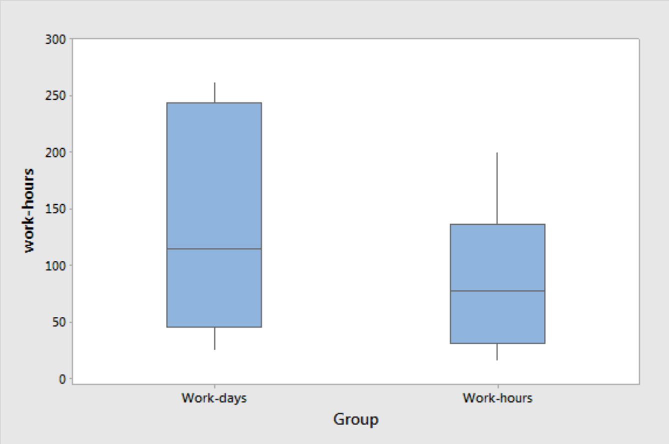

Fig. 3. Comparison of estimates in work-days and work-hours

It is not clear which of the two groups that had the most accurate effort estimates, only that the choice of effort unit affected the participants' judgments about the effort estimates. A slightly extended version of the project was completed by seven companies, see [23], with actual effort ranging from about 50 to 500 work-hours with a median of about 200 work-hours. Comparing with these numbers, it may seem as if those in the work-hours group tended to under-estimate and that those in the work-days group were the most realistic. This is, however, difficult to say without knowing how much the teams assumed by the participants in this experiments would actually use.

V. Conclusions and Further Work

In two experiments we have found a strong effect on the estimation unit on the resulting estimates. Estimates in minutes gave higher estimates than estimates in seconds, and estimates in work-days gave higher estimates than estimates in work-hours. The two main mechanisms responsible for this effect is, we argue, the numerosity and the unitosity effect.

In a situation with a tendency towards too low effort estimates, such as software development effort estimation, the request for estimates in workdays instead of work-hours is consequently likely to reduce the over-optimism of the effort estimates.

Of importance, but not much addressed in our experiments, is the selection of the time unit that gives the most accurate estimates. It is possible that the fluency of a time unit is of importance for that purpose, i.e., that it is essential to select the unit that is most natural for the work at hand. For very large projects this may mean that work packages would better be estimated in man-months or man-weeks, while for smaller tasks work-hours would be the better choice. More studies should be conducted to examine this topic, as well as studies on the effect sizes of choice of effort unit on effort estimates in real-world contexts.

References

1. Moløkken, K., M. Jørgensen, and I.C.S. Ieee Computer Society. *A review of surveys on software effort estimation*. in *International Symposium on Empirical Software Engineering*. 2003. Rome, Italy: Ieee Computer Soc.
2. Flyvbjerg, B. and A. Budzier, *Why your IT project may be riskier than you think.* Harvard Business Review, 2011. **89**(9): p. 601-603.
3. Jorgensen, M., *What We Do and Don't Know about Software Development Effort Estimation.* Software, IEEE, 2014. **31**(2): p. 37-40.
4. Halkjelsvik, T. and M. Jørgensen, *From origami to software development: A review of studies on judgment-based predictions of performance time.* Psychological Bulletin, 2012. **138**(2): p. 238-271.
5. Irving, F., *The money illusion.* New York. Adelphi, 1928.
6. Cannon, E.S. and G.P. Cipriani, *Euro-illusion: A natural experiment.* Journal of Money, Credit, and Banking, 2006. **38**(5): p. 1391-1403.
7. Ülkümen, G. and M. Thomas, *Personal Relevance and Mental Simulation Amplify the Duration Framing Effect.* Journal of Marketing Research, 2013. **50**(2): p. 194-206.
8. Ülkümen, G., M. Thomas, and V.G. Morwitz, *Will I spend more in 12 months or a year? The effect of ease of estimation and confidence on budget estimates.* Journal of Consumer Research, 2008. **35**(2): p. 245-256.
9. Monga, A. and R. Bagchi, *Years, months, and days versus 1, 12, and 365: The influence of units versus numbers.* Journal of Consumer Research, 2012. **39**(1): p. 185-198.
10. Grice, H., *Logic and Conversation'in P. Cole and J. Morgan (eds.) Syntax and Semantics Volume 3: Speech Acts*. 1975, Academic Press, New York.
11. Løhre, E. and M. Jørgensen, *Numerical anchors and their strong effects on software development effort estimates.* submitted for publication, 2014.
12. Jørgensen, M. and S. Grimstad, *Avoiding irrelevant and misleading information when estimating development effort.* IEEE Software, 2008. **25**(3): p. 78-83.
13. Jørgensen, M. and D.I.K. Sjøberg, *The impact of customer expectation on software development effort estimates.* International Journal of Project Management, 2004. **22**: p. 317-325.
14. Trope, Y. and N. Liberman, *Construal-level theory of psychological distance.* Psychological review, 2010. **117**(2): p. 440.
15. Liberman, N. and Y. Trope, *The role of feasibility and desirability considerations in near and distant future decisions: A test of temporal construal theory.*

Journal of personality and social psychology, 1998. **75**(1): p. 5.
16. Maglio, S.J. and Y. Trope, *Scale and construal: How larger measurement units shrink length estimates and expand mental horizons.* Psychonomic bulletin & review, 2011. **18**(1): p. 165-170.
17. Kanten, A.B., *The effect of construal level on predictions of task duration.* Journal of Experimental Social Psychology, 2011. **47**(6): p. 1037-1047.
18. Yan, J., S. Hou, and A. Unger, *High construal level reduces overoptimistic performance prediction.* Social Behavior and Personality: an international journal, 2014. **42**(8): p. 1303-1313.
19. Siddiqui, R.A., F. May, and A. Monga, *Reversals of task duration estimates: Thinking how rather than why shrinks duration estimates for simple tasks, but elongates estimates for complex tasks.* Journal of Experimental Social Psychology, 2014. **50**: p. 184-189.
20. Oppenheimer, D.M., *The secret life of fluency.* Trends in cognitive sciences, 2008. **12**(6): p. 237-241.
21. Song, H. and N. Schwarz, *If it's hard to read, it's hard to do processing fluency affects effort prediction and motivation.* Psychological Science, 2008. **19**(10): p. 986-988.
22. Jorgensen, M. and S. Grimstad, *The Impact of Irrelevant and Misleading Information on Software Development Effort Estimates: A Randomized Controlled Field Experiment.* Ieee Transactions on Software Engineering, 2011. **37**(5): p. 695-707.
23. Jørgensen, M., *Better Selection of Software Providers Through Trialsourcing.* To appear in IEEE Software, 2015.