Five things you should know about cost overrun

Flyvbjerg, Bent; Ansar, Atif; Budzier, Alexander; Buhl, Søren; Cantarelli, Chantal; Garbuio, Massimo; Glenting, Carsten; Holm, Mette Skamris; Lovallo, Dan; Lunn, Daniel

DOI
10.1016/j.tra.2018.07.013

Publication date
2018

Document Version
Final published version

Published in
Transportation Research Part A: Policy and Practice

Citation (APA)
Flyvbjerg, B., Ansar, A., Budzier, A., Buhl, S., Cantarelli, C., Garbuio, M., Glenting, C., Holm, M. S., Lovallo, D., Lunn, D., Molin, E., Rønnest, A., Stewart, A., & van Wee, B. (2018). Five things you should know about cost overrun. *Transportation Research Part A: Policy and Practice*, *118*, 174-190. https://doi.org/10.1016/j.tra.2018.07.013

Transportation Research Part A 118 (2018) 174–190

Contents lists available at ScienceDirect

Transportation Research Part A

journal homepage: www.elsevier.com/locate/tra

Five things you should know about cost overrun☆

Bent Flyvbjerg[a,*], Atif Ansar[a], Alexander Budzier[a], Søren Buhl[b], Chantal Cantarelli[c], Massimo Garbuio[d], Carsten Glenting[e], Mette Skamris Holm[f], Dan Lovallo[d], Daniel Lunn[g], Eric Molin[h], Arne Rønnest[i], Allison Stewart[j,1], Bert van Wee[h]

[a] The University of Oxford's Saïd Business School, United Kingdom
[b] Aalborg University's Department of Mathematical Sciences, Denmark
[c] Sheffield University Management School, United Kingdom
[d] University of Sydney Business School, Australia
[e] Viegand Maagøe A/S, Copenhagen, Denmark
[f] Aalborg Municipality, Denmark
[g] University of Oxford's Department of Statistics, United Kingdom
[h] Delft University of Technology's Faculty for Technology, Policy, and Management, Netherlands
[i] Esrum Kloster and Møllegård, Græsted, Denmark
[j] Infrastructure Victoria, Melbourne, Australia

ARTICLE INFO

Keywords:
Cost overrun
Cost underestimation
Cost forecasting
Root causes of cost overrun
Behavioral science
Optimism bias
Strategic misrepresentation
Delusion
Deception
Moral hazard
Agency
Reference class forecasting
De-biasing

ABSTRACT

This paper gives an overview of good and bad practice for understanding and curbing cost overrun in large capital investment projects, with a critique of Love and Ahiaga-Dagbui (2018) as point of departure. Good practice entails: (a) Consistent definition and measurement of overrun; in contrast to mixing inconsistent baselines, price levels, etc. (b) Data collection that includes all valid and reliable data; as opposed to including idiosyncratically sampled data, data with removed outliers, non-valid data from consultancies, etc. (c) Recognition that cost overrun is systemically fat-tailed; in contrast to understanding overrun in terms of error and randomness. (d) Acknowledgment that the root cause of cost overrun is behavioral bias; in contrast to explanations in terms of scope changes, complexity, etc. (e) De-biasing cost estimates with reference class forecasting or similar methods based in behavioral science; as opposed to conventional methods of estimation, with their century-long track record of inaccuracy and systemic bias. Bad practice is characterized by violating at least one of these five points. Love and Ahiaga-Dagbui violate all five. In so doing, they produce an exceptionally useful and comprehensive catalog of the many pitfalls that exist, and must be avoided, for properly understanding and curbing cost overrun.

1. Five key questions about cost overrun

Cost overrun in large capital investment projects can be hugely damaging, incurring outsize losses on investors and tax payers, compromising chief executives and their organizations, and even leading to bankruptcy (Flyvbjerg et al., 2009; Flyvbjerg and

☆ All authors have co-authored or authored publications based on the data, theories, and methods commented on by Love and Ahiaga-Dagbui (2018).

* Corresponding author.
E-mail address: bent.flyvbjerg@sbs.ox.ac.uk (B. Flyvbjerg).

[1] Former Associate Fellow at University of Oxford's Saïd Business School.

https://doi.org/10.1016/j.tra.2018.07.013
Received 19 July 2018; Accepted 19 July 2018
Available online 12 September 2018
0965-8564/

Budzier, 2011). Accordingly, cost overrun receives substantial attention in both the professional literature and popular media. Yet it is not always clear how cost overrun is defined, why it happens, and how to best avoid it, which has led to misperceptions about the concept with policy makers, planners, investors, academics, and the public. To help remedy this situation, below we address five fundamental questions about cost overrun in large capital investment projects:

1. What is cost overrun, and how is it measured?
2. Which data are used to establish cost overrun?
3. What is the size and frequency of cost overrun?
4. What are the root causes of cost overrun?
5. How is cost overrun best avoided?

If your job is to research, plan, finance, or deliver large capital projects, you need to have good answers to these questions. Here, we answer the questions in a response to Love and Ahiaga-Dagbui (2018), invited by the editors. We appreciate this opportunity to clarify what good and bad practice is in understanding and curbing cost overrun, and the many pitfalls that exist for good practice, eminently exemplified by Love and Ahiaga-Dagbui.

We are delighted that Love and Ahiaga-Dagbui acknowledge that our work on cost underestimation, "Undeniably … has made an impact … [and] brought to attention issues that were possibly being overlooked … The 'elephant in the room' has been recognized," as they say (p. 359). To understand and deal with the "elephant in the room" – deliberate and non-deliberate cost underestimation in large capital investment projects – has been a core purpose of our work. For Love and Ahiaga-Dagbui to recognize that we have succeeded is gratifying, and we thank them for their acknowledgment.

Love and Ahiaga-Dagbui, however, are critical of our work. We welcome their objections, as criticism is the main mechanism for securing high levels of validity and reliability in scholarship. But we are surprised by the language used by Love and Ahiaga-Dagbui in communicating their commentary. For instance, they describe our research findings as "fake news", "myths" (no less than 15 times), "canards", "factoids", "flagrant", "rhetoric", "misinformation", and more. We are further accused of having "fooled many people" by having "been just as crafty as Machiavelli" as we "have feigned and dissembled information" through our research (p. 358). As a factual observation, in our entire careers we have never come across language in an academic journal like that used by Love and Ahiaga-Dagbui. We suggest such language has no place in academic discourse.

In what follows, we address Love and Ahiaga-Dagbui's critique by relating it to each of the five key questions about cost overrun listed above.

2. What is cost overrun, and how is it measured?

Cost overrun is the amount by which actual cost exceeds estimated cost, with cost measured in the local currency, constant prices, and against a consistent baseline. Overrun is typically measured in percent of estimated cost, with a positive value indicating cost overrun and a negative value underrun. Size, frequency, and distribution of cost overrun should all be measured as part of measuring cost overrun for a certain investment type.

Cost overrun is the difference between actual and estimated capital costs for an investment. The difference may be measured in absolute or relative terms. In absolute terms cost overrun is measured as actual minus estimated cost.[2] In relative terms overrun is measured as either (a) actual cost in percent of estimated cost, or (b) the ratio of actual divided by estimated cost. In our studies, we measure cost overrun in relative terms, because this makes for accurate comparison across investments, geographies, and time periods. It also makes for accuracy in forecasts of cost risk.

Estimated cost may be, and typically is, established at different time points – or baselines – in the investment and delivery cycle, e.g., at the outline business case, final business case, and contracting. The cost estimate will normally be different at different time points, and it typically becomes more accurate the closer to final delivery the investment is, although there may be large variations in this, for instance where bad news about cost overruns are hidden as long as possible and cost estimates therefore suddenly explode when the project is well into delivery, when the overruns can no longer be kept secret, which is not an uncommon occurrence for large capital investment projects.

The baseline one chooses for measuring cost overrun depends on what one wants to understand and measure. We want to understand whether decision makers make well-informed decisions. For cost, this means we want to know whether the cost estimate on the basis of which decision makers decide to go ahead with a project is accurate. If the cost estimate is accurate the decision makers were well informed; if the estimate is inaccurate they were ill informed. We therefore use this cost estimate – called *the budget at the time of decision to build* – as baseline for measuring cost overrun in our studies, including Flyvbjerg et al. (2002, 2004), Cantarelli et al. (2010a; 2012a,b,c), Flyvbjerg et al. (2009), Flyvbjerg and Budzier (2011), Flyvbjerg (2014a, 2014b), Ansar et al. (2016, 2014),

[2] Actual cost is defined as real, accounted capital investment costs determined at the time of completion of the investment, when expenditures are known. Estimated cost is defined as budgeted, or forecasted, capital investment cost. Actual and estimated cost are calculated in the local currency and at the same price level (e.g., 2017 dollars) to ensure their comparability. Typically financing cost (e.g., interest payment on loans) is not included in estimated and actual capital investment cost.

Flyvbjerg (2016), and Flyvbjerg et al. (2016).[3] Love and Ahiaga-Dagbui's critique applies to these publications because they all use the data, theory, and methodology first set out in Flyvbjerg et al. (2002), which is the main focus of Love and Ahiaga-Dagbui although they also cite several of the other publications.

Love and Ahiaga-Dagbui object to this baseline. "The use of the budget at the decision-to-build may lead to inflated cost overruns being propagated," they claim (p. 363). They recommend to use the budget at contracting as baseline instead, which would on average show a lower cost overrun because this baseline is placed later in the investment cycle. We maintain that *the choice of baseline should reflect what it is you want to measure*, and discuss this in detail in Cantarelli et al. (2010a, 2010b). We agree that if you want to measure the performance of contractors, e.g., how well they deliver to the contracted budget, which seems to be Love and Ahiaga-Dagbui's focus, then the baseline at contracting is the right choice.[4] However, this is not what we wanted to measure in our studies. Our focus is on decision making, and hence on measuring the accuracy of information available to decision makers. With this focus, the budget at the time of making the decision to build is the right baseline for measuring cost overrun. Without this baseline it would be impossible to answer the important question of whether decisions are well informed or not regarding cost. Budgets estimated after the decision to build – like the contracted budget – are by definition irrelevant to this decision. Whatever the reasons are for inaccurate budgets, if the aim is to improve the quality of decision making then legislators, citizens, investors, planners, and managers must know the uncertainty of budget forecasts, which is what we measure in our studies. Otherwise transparency, accountability, and good governance will suffer.

We agree with Love and Ahiaga-Dagbui that cost estimates will vary, "depending on governments' decision-making processes" and that, "This makes it difficult to understand and compare project costs between markets" (p. 361). However, we do not compare costs between markets, we compare *percentage cost overrun*. Love and Ahiaga-Dagbui's text indicate they do not understand the difference between comparing cost and comparing cost overrun, i.e., comparing an absolute versus comparing a relative value, which is fundamentally different in statistical terms, with comparison of relative values (percentage cost overrun) possible and desirable.[5]

We further agree with Love and Ahiaga-Dagbui that different technologies, funding mechanisms, legal systems, and environmental regulations produce substantial variations across jurisdictions in the cost estimating process and in cost estimates. We disagree, however, that this means "that undertaking any form of comparative study on the accuracy of estimated costs with this [Flyvbjerg et al.'s] dataset would be nonsensical" (p. 361). If the statistical model is properly set up, whatever factors are of interest can be compared, and the differences present opportunities to study the influence on cost overrun of dissimilarities in technology, funding mechanisms, etc. The differences are possible explanations of differences in cost overrun, and we study them as such, e.g., in Flyvbjerg et al. (2004) and Cantarelli and Flyvbjerg (2015), as do others (Pickrell, 1990; Fouracre et al., 1990; Terrill et al., 2016; Welde and Odeck, 2017).

Finally, we agree with Love and Ahiaga-Dagbui that "no international standards exist to determine the level of detail needed to formulate an estimate at the time the decision-to-build is made" (pp. 360–61). An international standard does exist, however, to use the budget at the time of decision to build as baseline in studies that specifically seek to understand how well informed decision makers are in deciding whether to invest or not in large capital projects. This is the standard we refer to and use in our studies. It was originally laid down in Pickrell (1990), Fouracre et al. (1990), Walmsley and Pickett (1992), National Audit Office (1992), Leavitt et al. (1993), World Bank (1994), Nijkamp and Ubbels (1999), and Flyvbjerg et al. (2002). The standard has since been followed in dozens of other studies and is used by government in, e.g., the USA, China, the UK, Scandinavia, and The Netherlands to assess and estimate cost overrun. But nowhere do we claim that this is a general standard, because other baselines may be used for other purposes, as explained above, something Love and Ahiaga-Dagbui seem to overlook. Because we do not claim that the decision to build is a general standard, Love and Ahiaga-Dagbui's Myth No. 2 (that we do) falls flat. It is irrelevant (see Table 1).

3. Which data are used to establish cost overrun?

When establishing cost overrun for a specific project type, typically a sample from a population of projects is used. The sample must properly reflect the population of projects for which one wants to measure cost overrun. All relevant projects, for which valid and reliable data are available, should be included in the sample. Any bias in the sample must be carefully considered.

In statistical analysis, data are ideally a sample drawn from a larger population, and the sample represents the population properly. These requirements may be satisfied by drawing the sample by randomized lot. In studies of human affairs, however, in

[3] To be more specific, we baseline cost overrun in the budget at the time of the *formal* decision to build, i.e., the time the formal decision making body (e.g., parliament, government authority, board of directors) approved a project to go ahead. Often the *real* decision to build a project has been made much earlier, with a point of no return. Measured against the budget at the time of the real decision to build, cost overrun is typically substantially higher than when measured against the budget at the time of the formal decision to build, introducing a conservative bias in our findings. We cover this problematic in Cantarelli et al. (2010b).

[4] Similarly, if you wanted to understand how accurate cost estimates are in outline business cases, you would use the outline business case as baseline. Or if you wanted to understand half way through construction of a project how likely the remaining budget is to be sufficient, you would baseline here, as Flyvbjerg et al. (2016) did for Hong Kong's XRL high-speed rail project. The baseline depends on what you want to understand. There is not one given baseline that is right, as Love and Ahiaga-Dagbui seem to think.

[5] This point also mutes Love and Ahiaga-Dagbui's critique that, "the use of currency conversations [sic, we assume Love and Ahiaga-Dagbui mean "conversions"] has no meaning for comparing the costs of construction unless issues associated with an economy's Purchasing Power Parity are taken into consideration" (p. 365). Issues of purchasing power parity are relevant for comparing cost, we agree, but irrelevant for comparing percentage cost overrun, which is what we do.

Table 1
Reality check for Love and Ahiaga-Dagbui's four postulated myths.

Postulated myth	Reality
Myth No. 1: The sample is the largest of its kind and the first statistically significant study of cost escalation.	Flyvbjerg et al. (2002) claimed their sample was the largest of its kind, i.e., the largest of academic datasets aimed at understanding cost overrun in the worldwide population of large transportation infrastructure projects. Love and Ahiaga-Dagbui present five studies of small US highway projects as evidence that larger samples with statistical analyses existed. However, three of the five studies did not exist when Flyvbjerg et al. (2002) was written; one of the remaining two has a smaller sample. Moreover, none of the five studies are comparable to Flyvbjerg et al. (2002), because (a) they are not aimed at understanding cost overrun in the worldwide population of large transportation infrastructure projects, (b) they are baselined differently, and (c) their data are not valid. Flyvbjerg et al.'s claim therefore stands
Myth No. 2: Using the budget at decision to build as baseline to determine cost underestimation is an international standard.	We do not claim that using the decision to build as baseline is a general standard. Love and Ahiaga-Dagbui's Myth No. 2 (that we do) therefore falls. Choice of baseline depends on what you want to measure. If, like us, you want to measure how well informed decision makers are when deciding to invest in large capital projects, then the decision to build is the right baseline. If, like Love and Ahiaga-Dagbui, you want to measure how well contractors deliver to their bid, then contracting is the right baseline. One baseline does not fit all purposes, as Love and Ahiaga-Dagbui seem to think
Myth No. 3: Costs are underestimated in nine out of 10 transportation infrastructure projects.	Love and Ahiaga-Dagbui present no valid and reliable data to counter Flyvbjerg et al.'s (2002) finding that 86 percent of large transportation infrastructure investments have cost overrun. Other studies confirm this finding, including one by Peter Love himself, which found a frequency of cost overrun of 95 percent (Love et al., 2012: 147). Love and Ahiaga-Dagbui's Myth No. 3 (that our finding is a myth) therefore falls and our finding stands
Myth No. 4: Underestimation cannot be explained by error and is best explained by strategic misrepresentation, that is, lying.	We maintain that cost overrun is best explained in terms of psychological and political biases with planners. Love and Ahiaga-Dagbui explain cost overrun in terms of errors caused by scope changes and complexity. We have on our side (a) Nobel-Prize-winning theory on heuristics and biases, (b) principal-agent theory, (c) high-quality data that fit the two sets of theory at an overwhelmingly high level of statistical significance ($p < 0.001$), and (d) planners stating on the record that they deliberately underestimated costs to have projects approved and funded, establishing the causal link between deception and cost underestimation that Love and Ahiaga-Dagbui claim does not exist. In comparison, Love and Ahiaga-Dagbui offer their untested "evolutionist" theory that (a) does not fit the data, because according to the theory inaccuracies in cost estimates would be random, but in reality they are systemically biased, and (b) has resulted in unsuccessful mitigation and containment strategies, according to Love and Ahiaga-Dagbui themselves. As the evidence stands, Love and Ahiaga-Dagbui's Myth No. 4 (that explanations in terms of bias are a myth) falls and our explanation stands, because it is better supported both theoretically and empirically

which controlled laboratory experiments often cannot be conducted, it is frequently impossible to meet these ideal conditions. So too for studies of cost overrun. One therefore has to take a different approach to sampling and statistical analysis in such studies. *Good practice is to include all relevant capital projects for which valid and reliable data are available, so that no distributional information goes to waste*. The latter is a crucial precondition for producing high-quality results, as argued by Kahneman (2011: 251), Lovallo and Kahneman (2003), Lovallo et al. (2012), and Flyvbjerg (2008, 2013a). Thus, the sampling criterion for good practice is data availability. However, this means that the sample will often not be representative of the population it is drawn from, because availability does not necessarily equate representativeness. It is therefore critical, as part of good practice, to *carefully assess any biases arising from sampling* in this manner. This is the approach taken to data collection in our research, from the first to the most recent studies, carefully described in each publication, including in Flyvbjerg et al. (2002: 294–295).

Love and Ahiaga-Dagbui write that the original sample of 258 transportation infrastructure projects in Flyvbjerg et al. (2002) – later included in larger datasets in Flyvbjerg et al. (2004), Cantarelli et al. (2012a,b,c), Ansar et al. (2014), and Flyvbjerg (2016) – "falls short of providing a statistically representative sample of the population" (p. 361).[6] But the lack of representativeness is something we ourselves pointed out in the original study, in which – following the approach outlined above – we explicitly state: "the question is whether the projects included in the sample are representative of the population of transportation infrastructure projects … There are four reasons why this is probably not the case" (Flyvbjerg et al., 2002: 294).[7] Love and Ahiaga-Dagbui's restatement 16 years later is no contribution. We further established, in the original study and repeated it in later studies, that the lack of

[6] The original dataset in Flyvbjerg et al. (2002) was collected by Mette Skamris Holm and Bent Flyvbjerg; the larger dataset in Flyvbjerg et al. (2004) by Bent Flyvbjerg, Carsten Glenting, and Arne Rønnest; the dataset in Cantarelli et al. (2012a,b,c) by Chantal Cantarelli and Bent Flyvbjerg; the dataset in Ansar et al. (2016) by Atif Ansar, Bent Flyvbjerg, and Alexander Budzier; and the dataset in Flyvbjerg (2016) by Bent Flyvbjerg, Alexander Budzier, Chantal Cantarelli, and Atif Ansar. Love and Ahiaga-Dagbui (p. 360) say that ideally data should be made publicly available to allow for evaluation and reproduction of the results. We agree, but in practice sources often insist on non-disclosure agreements as a condition for making their data available for study.

[7] We repeated this in later studies, e.g., Cantarelli et al. (2012a: 3): "the current sample of 806 projects is probably not representative of the population of transport infrastructure projects … The sample is biased and the bias is conservative."

Table 2
Overview of statistical errors in Love and Ahiaga-Dagbui (2018). Each error is documented and explained in the main text.

No.	Statistical error
1	Love and Ahiaga-Dagbui attempt to calculate sample size based on population size, but population size does not matter when the population is considerably larger than the sample. They get the sample size wrong.
2	They say our sample size is too small, despite the fact that we establish overwhelmingly high levels of statistical significance for our main results ($p < 0.001$). Saying the sample size is too small when significance has been established shows a fundamental misunderstanding of statistics.
3	They criticize statistical meta-analysis for "piggy-backing" off data collected by others and for "over-reliance on secondary sources," which shows a cardinal misunderstanding of what statistical meta-analysis is. Best practice in meta-analysis is to use *all* available data, including data from other studies.
4	They confuse the sample with the universe from which the sample was drawn, thus misinterpreting the sample size in Thurgood et al. (1990) as being 817 when in fact it is 106.
5	They erroneously compare studies aimed at understanding fundamentally different populations, demonstrating lack of understanding of what a statistical population is and which populations may be compared.
6	They erroneously claim that studies of small projects are directly comparable to studies of large projects.
7	Love and Ahiaga-Dagbui mix up and compare studies that arrive at their conclusions by studying samples of projects with studies that arrive at their conclusions by studying whole populations of projects.
8	They mix up and compare studies that use different baselines in measuring costs, especially the budget at contracting with the budget at the decision to build, committing the age-old error of comparing apples and oranges.
9	They compare our studies with studies that are fundamentally and obviously flawed, e.g., a study that leaves out 56 outliers, rendering this study meaningless, and studies only looking at large cost overruns.
10	Their mistaken comments on the use of normal distributions in statistical tests demonstrate ignorance of the Central Limit Theorem, which is key to understanding such tests.
11	They claim that the production of an estimate "demands knowledge of what will occur." This is a deterministic fallacy. It demonstrates a lack of understanding of uncertainty and statistical estimation. For statistical estimation you do not need to know what *will* occur. You need to know the odds of what *may* occur. With the right data statistical analysis gives you these odds.
12	They seem to regard the standard deviation as generally independent of the mean, when that is only so for the special case of normal distributions, which rarely applies to cost overrun.
13	They recommend measuring uncertainty by the standard deviation. But distributions of cost overrun for large infrastructure projects are fat-tailed, and for such distributions the standard deviation is not a good measure of uncertainty.
14	Love and Ahiaga-Dagbui talk about "fitting … data to the distribution," which makes no sense at all in statistics where data are immutable and distributions are fitted to the data, or a mathematical transformation of them, not the other way around.

representativeness means there is a conservative bias in our data, i.e., cost overrun in the project population is likely to be higher than cost overrun in the sample,[8] which Love and Ahiaga-Dagbui fail to mention.

Love and Ahiaga-Dagbui make an attempt at deciding sample size based on population size and they estimate that "the optimum sample size should have been 383," i.e., 125 projects more than our original sample, for the sample to be representative (p. 361). We respectfully disagree and see this as one of many statistical errors committed by Love and Ahiaga-Dagbui (for an overview of these errors, see Table 2). Population size does not matter, when the population is considerably larger than the sample size.[9] E.g., an opinion poll for Iceland, with a population of 350,000, requires the same sample size as one for the USA, with a population of 326 million. In fact, any sample size will do when significance has been shown, like in our studies, assuming the sampling and the calculations have been done properly. This means that our dataset has a size that is big enough to show what we have shown, with p-values less than 0.05, and even 0.001 and smaller for our main results. In sum, saying that the sample size is too small when significance has actually been established indicates a fundamental misunderstanding of what statistics is. We will see more errors like this below.

We agree with Love and Ahiaga-Dagbui, however, that larger samples are desirable for the study of cost underestimation and overrun. This is why we have continued to enlarge and update the original dataset, to 806 projects in Cantarelli et al. (2012a) and 2062 in Flyvbjerg (2016), now including both costs and benefits. The later studies show the original 2002 results to be robust across significantly more observations, more project types, more geographies, and a longer time span. So much for Love and Ahiaga-Dagbui's "debunking" of the 2002 study (pp. 358, 360): the original results have been replicated and confirmed with more and better data, which is a key scientific criterion for proving the validity and robustness of results. Even larger datasets are in the pipeline to test our conclusions further and to investigate new problematics. To be credible, Love and Ahiaga-Dagbui would have

[8] There are four reasons for this bias. First, that projects that are managed well with respect to data availability are likely to also be managed well in other respects, resulting in better than average performance. Second, it has been argued that the very existence of data that make the evaluation of performance possible may contribute to improved performance when such data are used by project management to monitor projects (World Bank 1994: 17). Third, managers of projects with a particularly bad track record regarding cost overrun may have an interest in not making cost data available, which would then result in underrepresentation of such projects in the sample. Conversely, managers of projects with a good track record for cost might be interested in making this public, resulting in overrepresentation of these projects. Fourth, and finally, even where managers have made cost data available, they may have chosen to give out data that present their projects in as favorable a light as possible. Often there are several estimates of costs to choose from and several calculations of actual costs for a given project at a given time. Managers may decide to choose the combination of actual and estimated costs that suits them best, possibly a combination that makes their projects look good with a low cost overrun.

[9] This and any other statement about statistics in the present and previous papers have been formulated or verified by professional mathematical statisticians, who form part of the author team for our papers, including the present paper.

had to present datasets and studies showing different results from ours. They do not. Instead they postulate their debunking of our work.

We further contest Love and Ahiaga-Dagbui on the following points about the dataset. First, Love and Ahiaga-Dagbui claim we have been "cherry-picking data" (p. 357). This is a serious allegation. It would imply fraud on our part. But Love and Ahiaga-Dagbui provide no evidence to support the accusation. Moreover, the charge is easy to disprove, because from 2002 until today our sampling criterion has always been, and was always explicitly stated as: "all projects for which data were considered valid and reliable were included in the sample" (Flyvbjerg et al., 2002: 294, Cantarelli et al., 2012a,b,c). "All" means *all*, so no cherry-picking would be possible without violating this criterion, which would also violate Kahneman's advice of including all distributional information. We take this piece of advice seriously, because it is key to the quality of our data and the accuracy of our forecasts. The burden of proof is on Love and Ahiaga-Dagbui to show where we violated our sampling criterion and cherry-picked data. If Love and Ahiaga-Dagbui cannot show this, their accusation falls and all that is left is their loud rhetoric, here as elsewhere in their paper.

Second, according to Love and Ahiaga-Dagbui, we have "piggy-backed off the data collected and published by other researchers" (p. 360). This statement is another statistical error on the part of Love and Ahiaga-Dagbui. In statistics, using data from other studies is called meta-analysis, not piggy-backing, and it is considered best practice. Much of statistics would be impossible if combination of data from different sources was forbidden. We did meta-analysis, because it would be an error not to, as this would have left out relevant information from our studies (i.e., we would have cherry-picked).[10] We explicitly acknowledge and cite the studies we draw on with full references (Flyvbjerg et al., 2002: 294). In short, we chose to stand on the shoulders of our colleagues to make a cumulative contribution to our field. What do Love and Ahiaga-Dagbui suggest as good practice, if not this? They do not answer this question. To be credible they would have to.

Third, Love and Ahiaga-Dagbui claim we have an "over-reliance on secondary sources" and that this "reaffirms the unreliable nature of the data" (p. 362). This is a further statistical error. In statistical meta-analysis one does not speak of "secondary sources," but of "other studies." Furthermore, to produce the most valid and reliable results one has to include *all* other studies with valid and reliable data. It is non-sensical to speak of "over-reliance." Sometimes "other studies" account for 100 percent of observations in meta-analyses, and even then no statistician would speak of "over-reliance." Finally, there is nothing unreliable about the studies we included. Others and we have established their reliability before we included them in our study, as is standard for meta-analyses.

Fourth, we stated in Flyvbjerg et al. (2002: 280, 293) and later papers (Cantarelli et al., 2012a,b,c; Flyvbjerg, 2016; Ansar et al., 2014) that, as far as we know, our dataset on cost overrun in large transportation infrastructure projects is the largest of its kind. We explained in these publications that this means largest among independent datasets that have been subject to academic peer review and can be considered to contain valid and reliable data suitable for scholarly research aimed at assessing cost overrun in the global population of large transportation infrastructure projects. Love and Ahiaga-Dagbui question our claim: "More reliable statistical studies using larger samples had been undertaken prior to Flyvbjerg et al. (2002) study" (p. 362). As examples of such studies, Love and Ahiaga-Dagbui cite five studies of small highway projects in four US states, each sponsored by a state department for transportation (Thurgood et al., 1990; Hinze and Selstead, 1991; Vidalis and Najafi, 2002; Bordat et al., 2004; Ellis et al., 2007).[11] Disregarding the fact that three of the five studies did not exist when Flyvbjerg et al. (2002) was written, and that one of the remaining two (Thurgood et al., 1990) in fact has a smaller sample, none of the five studies is comparable to Flyvbjerg et al. (2002) for the reasons listed below.

Love and Ahiaga-Dagbui claim that Thurgood et al. (1990), "examined the cost overruns of 817 highway projects delivered by the Utah Department of Transportation … between 1980 and 1989" (p. 362).[12] The authors misread Thurgood et al. (1990: 122), however, who explicitly write, "The final sample size was 106 projects," which is less than half the size of our smallest sample on cost overrun. Love and Ahiaga-Dagbui seem to confuse the universe that Thurgood et al. drew their sample from (which consists of 817 projects) with the sample itself, which is yet another statistical error. Measured in dollar-value, our sample is in fact 708 times larger than Thurgood et al.'s, because they focus on small projects, as we will see below, while we focus on large ones.

But this is not Love and Ahiaga-Dagbui's worst misunderstanding regarding the five studies. None of the studies is comparable to ours, irrespective of sample size. The studies are not even comparable to each other, as we will see. The comparison that Love and Ahiaga-Dagbui make is false for the following reasons:

(a) The five studies cover highway projects in each of four US states only.[13] Our studies cover highways, freeways, bridges, tunnels,

[10] Typically, we had access to the datasets we included in meta-analysis. We critically studied the data and research methodology behind each dataset to establish whether the data could be considered valid and reliable. Data that were considered not valid and reliable were rejected.

[11] We have earlier considered these studies, and more, for inclusion in our meta-analyses of cost overrun. Our standard procedure is that each time a new study is published we assess validity, reliability, and comparability of its data. If the data are valid, reliable, and comparable with ours (same baseline, constant prices, etc.) then we include the data for meta-analysis. If not, we reject the data. For the 2002 study, we considered all studies existing at the time. Below, we present what we found when we assessed validity, reliability, and comparability of the studies cited by Love and Ahiaga-Dagbui.

[12] The correct period is 1980 to 1988, not 1980 to 1989 as Love and Ahiaga-Dagbui wrongly claim.

[13] Thurgood et al. (1990) studied highways specifically with the Utah Department of Transportation, Hinze and Selstead (1991) the Washington State Department of Transportation, Vidalis and Najafi (2002) the Florida Department of Transportation, Bordat et al. (2004) the Indiana Department of Transportation, and Ellis et al. (2007) again the Florida Department of Transportation. The studies were sponsored by the individual departments of transportation with the purpose of curbing cost overrun within their respective jurisdictions.

urban heavy rail, urban light rail, conventional inter-city rail, and high-speed rail in 20 nations on five continents.[14] For this reason alone the five US studies are not comparable to our global study. None of the five studies aims to understand cost overrun in the global population of large transportation infrastructure projects, nor do they claim to, as opposed to Love and Ahiaga-Dagbui, who misrepresent the studies in making this claim when they say the studies are comparable to ours. Each of the five studies claims only to say something about the population of highway projects in the individual state studied. It is not surprising that specific geographies, like individual US states, have overruns that are different from global averages. It is to be expected and is also the case for specific geographies within our global sample, as we show in Cantarelli et al. (2012c), where we deal with heterogeneity and country bias. Love and Ahiaga-Dagbui make the error of comparing datasets that have been selected and studied with fundamentally different purposes.

(b) The five US studies cover only very small projects, each costing one or a few million dollars on average, whereas we study large transportation infrastructure projects with an average cost of $349 million.[15] To claim, without evidence, that studies of projects with an average size of one or a few million dollars are directly comparable to studies of projects costing hundreds of millions of dollars is a postulate with no credible foundation. Again, for this reason alone the five US studies are not comparable to our study.[16]

(c) Thurgood et al. (1990: 122) deliberately sampled for “unusually large overruns,” because they wanted to explain such overruns, with a view to helping their client, the Utah Department of Transportation, getting large overruns under control.[17] This means that Thurgood et al.'s sample is deliberately and openly biased, something Love and Ahiaga-Dagbui fail to mention, and something that again makes the sample not comparable with ours, or with other studies, including the other four US highways studies cited by Love and Ahiaga-Dagbui. It is misrepresentation, deliberate or not, on the part of Love and Ahiaga-Dagbui to fail to mention this bias and to present Thurgood et al.'s study as if it may be compared to other studies.[18]

(d) Similarly, Vidalis and Najafi (2002: 2) deliberately biased their sample by focusing only on “projects that had experienced cost and time overruns,” because this was the problem they and their client, the Florida Department of Transportation, wanted to solve. In comparison, our sample, and most other academic samples, contain projects with both overrun and underrun to ensure that all relevant distributional information is included. Again this difference means that sampling for the Vidalis and Najafi study is idiosyncratic and cannot be compared with other studies, including ours and all other studies mentioned in this paper. Love and Ahiaga-Dagbui fail to mention this important fact and thus misrepresent Vidalis and Najafi (2002).

(e) The baseline for measuring cost overrun in all five US highway studies is the budget at contracting. This baseline was chosen because the authors and their clients wanted to understand how to better control contract costs in each of the states that commissioned the studies. Our baseline is the budget at the decision to build, because we want to understand whether decision makers are well informed about cost when they give the green light to projects. Both baselines are legitimate, as mentioned, but they focus on very different problematics, again making the five studies not comparable with ours.[19]

(f) With the five studies, Love and Ahiaga-Dagbui mix studies that arrive at their conclusions by studying samples of projects (e.g.,

[14] This is for the 2002 study. Our later studies with larger samples (Cantarelli et al. 2012a,b,c; Flyvbjerg 2016; and Ansar et al., 2016) cover more project types and nations.

[15] The average project size of Thurgood et al. (1990: 122-123) is $1.2 million; Hinze and Selstead (1991: 27-18) $1 to 2 million, of which 67 percent of projects are smaller than $1 million; Vidalis and Najafi (2002: 2) $2.7 million; Bordat et al. (2004: 55) approximately $1 million, of which 76 percent are smaller than $1 million, and finally Ellis et al. (2007: 19) $4.8 million in average project size. (There are small differences in the price level in which the average costs are given in the five studies, but nothing that changes the main conclusion).

[16] We also include some smaller projects in our samples for comparison purposes, i.e., to study the effect of size on project performance (Flyvbjerg et al. 2004). This entails including project size as a covariate in statistical modeling and checking for dependence on other covariates, which might be different for different sizes of project. For Love and Ahiaga-Dagbui to directly compare large international projects with small US state projects without such statistical modeling demonstrates an inept understanding of statistics.

[17] In full, Thurgood et al. (1990: 122, emphasis added) stipulate that their final sample of 106 projects "included All *projects with unusually large overruns* ($100,000 or larger), All projects carried out by *contractors with an unusually high percentage of contracts with large overruns*, All projects supervised by *project engineers with an unusually high percentage of projects with large overruns*, and A 10 percent random sample of the remaining projects."

[18] The idiosyncratic sampling of Thurgood et al. (1990) demonstrate statistical ineptitude similar to that of Love and Ahiaga-Dagbui. It is impossible to attribute causes reliably with no normal projects acting as controls. Thurgood et al. should have included the whole range of projects in their analysis if they wanted to identify reasons why some projects have unusually large overruns. Their sampling is therefore not so such much biased as it is incompetent, and adding a 10 percent random sample of "remaining projects" does not solve the problem. This comment also pertains to Vidalis and Najafi (2002), covered under the next point.

[19] It should also be mentioned that some of the five studies, e.g., Bordat et al. (2004), study cost overrun on contracts instead of cost overrun on projects. The two are typically not the same, because a project is normally made up of several contracts. It is not always clear in the five studies chosen by Love and Ahiaga-Dagbui when contracts are used instead of projects, which undermines the validity, reliability, and comparability of results. Love and Ahiaga-Dagbui must be faulted for making no mention of this. In any case, it again makes the results not comparable with ours, which are based on the study of projects, not contracts.

Thurgood et al., 1990) with studies that arrive at their conclusions by studying whole populations of projects (e.g., Ellis et al., 2007), without making this clear to readers.[20] This is yet another statistical error. For valid and reliable conclusions, it would have been necessary for Love and Ahiaga-Dagbui to distinguish between the two types of study, because how analysis is done and results interpreted are different for studies of samples versus studies of whole populations. Love and Ahiaga-Dagbui again fail to mention this issue.

(g) Finally, none of the five studies are independent academic research, nor are they presented as such by their authors. The studies were all commissioned by individual state departments for transportation to understand and curb their specific problems with contractor cost overrun. The studies are therefore more akin to consulting, with the issues of validity, reliability, and comparability this raises. That is not a critique of the studies as such. They may have served their purposes. But the studies do not meet the criteria of academic research without independent verification, and such verification has not happened, to our knowledge. It is misrepresentation on the part of Love and Ahiaga-Dagbui to present the five studies as if they are directly comparable to scholarly research, including ours.

In sum, each of the points above show the five cited studies do not have the same goals as ours and are not comparable with ours. Taken together, the points form a devastating verdict on Love and Ahiaga-Dagbui's claim that the studies are relevant in the current context, including for comparing sample sizes and statistical analysis. Our claim stands that our dataset is the "largest of its kind," i.e., the largest of academic datasets aimed at understanding cost overrun in the worldwide population of large transportation infrastructure projects. Love and Ahiaga-Dagbui's postulated Myth No. 1 is therefore not a myth, but a fact (see Table 1).

4. What is the size and frequency of cost overrun?

Studies of cost overrun in large capital investment projects show: (a) Overrun is typically significantly more frequent than underrun; (b) Average and median overrun are typically positive and significantly different from zero; and (c) Average overrun is typically higher than median overrun, indicating fat upper tails.

Love and Ahiaga-Dagbui take issue with the results of our research in the following manner: "A never-ending factoid … is that 9 out of 10 transport projects worldwide experience cost overruns. Despite the unrepresentative nature of the sample, many academics have and continue to [sic] peddle this canard. For example, Shane et al." (p. 364). To be precise, our data show that 86 percent of transportation infrastructure projects have cost overrun, here rounded up to nine out of ten. Other studies confirm frequencies of this order of magnitude, some of them even higher, e.g., Dantata et al. (2006), Federal Transit Administration (2003, 2008), Lee (2008), Pickrell (1990), and Riksrevisionsverket (1994). We dealt with the representativeness of our sample above and argued that it is conservative, i.e., the real frequency of overrun in the project population is probably higher than the 86 percent found for the sample. Interestingly, Peter Love himself, in one of his main studies, found data to support this claim, with a frequency of cost overrun of 95 percent in a sample of 58 Australian transportation infrastructure projects (Love et al., 2012: 147).[21] On that background, we are at a loss to understand Love and Ahiaga-Dagbui's complaint about our 86 percent. To call this number a canard, as Love and Ahiaga-Dagbui do, is an insult to ducks and would require Love and Ahiaga-Dagbui to prove it wrong by presenting better data than ours, and different from Love's own data. Love and Ahiaga-Dagbui do no such thing, they simply postulate their critique in non-credible contrast to Love's own findings. To criticize Shane et al. on that basis is not only wrong, but unfair. Shane et al., like many others, simply cite the best available data in the research literature.

In an attempt to support their point, Love and Ahiaga-Dagbui quote a study from the Grattan Institute, an Australian think tank, which states that, "The majority of [Australian transportation infrastructure] projects come in close to their announced costs" (Terrill and Danks, 2016: 10). However, closer examination of this study and its data proves this to be an exceedingly selective and dubious quote, even if, theoretically, it is possible, of course, that Australian infrastructure projects differ significantly from the rest of the world.

First, Terrill and Danks (2016:11) make the following unusual assumption about the majority of projects they studied:

> "For the 68 per cent of projects where data on their early costs is missing in our dataset, we have made the assumption that no early cost overruns occurred."[22]

[20] For instance, Ellis et al. (2007: 17) studied populations of projects, which they describe in the following manner: "All completed projects from January 1998 to March 2006 were included in the search, which totaled 3130 projects. Of this total, 1160 projects were alternative contracting projects. The collected project data was initially reviewed to identify any obviously erroneous data. After the data cleaning process, a total of 3040 FDOT construction project data were used for analysis. The final dataset consisted of 1132 projects using alternative contracting methods and 1908 projects using traditional design-bid-build contracting methods." In contrast, Thurgood et al. (1990) studied a sample of projects, as described above.

[21] Figure 2 in Love et al. (2012: 147) shows that three of 58 projects in their sample have cost underrun. The remaining 55 projects, equivalent to 95 percent, have cost overrun. It should be mentioned that overrun for these projects was baselined at contracting, which, other things being equal, should result in a lower frequency of overrun than found in our studies, which are baselined at the decision to build, i.e., earlier in the project cycle, leaving more time for projects to accumulate overrun. This makes Love et al.'s finding of 95 percent conservative, compared to our finding, lending further support to our claim that overruns by far outnumber underruns.

[22] The Grattan Institute study suffers from several similar assumptions and shortcuts that negatively affect the quality of data and analyses. The explanation may be that the study was rushed. It was carried out in just a few months, which is short for a study of this magnitude, with all the quality control of data points and analyses that are needed for producing valid and reliable findings.

This is idiosyncratic sampling again. Our data indicate that for many projects the largest cost overruns occur in the early stages of the project cycle, so to assume zero cost overrun here is to assume away a major part of overrun (Cantarelli et al., 2012b). In our dataset, projects with such missing data are excluded to ensure valid and reliable results, which means that our data and results are not comparable to those of Terrill and Danks, who have the good grace to mention that their assumption means that the cost overruns in their report "may well be understated" and, "Cost overruns may be even bigger than we claim" (p. 11). Love and Ahiaga-Dagbui, engaging once more in misrepresentation, make no mention of this admission.

Second, if the Grattan Institute researchers found no public evidence of a cost overrun on a project they assumed that the project came in on budget. Their report hereby commit the fallacy of assuming that absence of evidence is evidence of absence, a no go for anyone who understands statistics. We do not accept absence of evidence as evidence in our studies. If we do not find evidence that a project actually came in on budget we do not include the project as a data point. The Grattan Institute study may therefore not be compared to our studies, or, indeed, any other study.

Third, Terrill and Danks main measure of cost overrun is baselined "after principal contracts have been awarded" (Terrill et al., 2016: 8). We measure cost overrun from the date of decision to build, which for large projects may be years earlier than contract award. Both baselines are legitimate, as said, depending on what you want to measure, but the different baselines mean that our results are not comparable with those of Terrill and Danks.

Fourth, Terrill and Danks use two different datasets in their analysis. One consists of 51 projects for which they themselves manually collected the data. According to this dataset, 65 percent of projects had cost overrun (compared with 86 percent for our data) with an average cost overrun of 52 percent (compared with 28 percent for our data). Thus, using their own data, the Grattan Institute study found a larger average cost overrun than us and indication of a clear skew in overrun, like us, something one would never know from reading Love and Ahiaga-Dagbui's selective use of results from the study.

Fifth, the other dataset used by Terrill and Danks consists of 542 projects from a database owned by a consultancy, from which data were collected electronically (Terrill et al., 2016: 5, 8). The majority of Terrill and Danks's findings are based on this second dataset, according to which 34 percent of projects had cost overrun and average cost overrun was 24 percent. From long experience working in and with consultancies, we know they typically do not have the time it takes to establish the level of validity and reliability required for peer-reviewed scholarly research, nor is this their purpose. Sometimes consultancies also have a tendency to find what their clients like to see, because it is good for business.[23] For this reason alone, we would reject the consultancy data, given that they have not been independently verified.

Sixth, and perhaps most importantly, Terrill and Danks (2016: 64) decided to remove 56 outliers from the consultancy dataset in an arbitrary and unaccounted for manner. Removal of outliers is justified only if they come from a different population of uncertainty than non-outlying observations, for instance due to erroneous measurements or experimenter bias. Blanket removal of outliers is a mistake – and a big one, when outliers are big as they are for cost overrun in capital investment projects (Steele and Huber, 2004). Removal of outliers before analysis is like trying to understand how best to earthquake proof buildings without taking the biggest earthquakes into account – not a good idea. Specifically, Terrill and Danks removed projects from the consultancy dataset if they had a cost overrun higher than the highest overrun in the smaller dataset of 51 projects mentioned above. This is an unusual way to define and eliminate outliers, which will not be found in any statistics textbook. Moreover, 56 outliers out of a total of 542 observations – more than ten percent – is simply not credible as a measurement error. Terrill and Danks did not so much remove outliers, as they artificially removed a real skew in their data. By so doing, they artificially lowered the frequency, average, median, and variance of cost overrun, which is exactly the opposite of what you want to do if your aim is valid and reliable data and analyses. Love and Ahiaga-Dagbui do not mention this, again misrepresenting the study.

When the Grattan Institute study came out in 2016, like for all such studies, we reviewed the data for possible inclusion in our meta-analyses. However, after carefully assessing the dataset, including discussions with Grattan Institute staff, we decided against inclusion, for the reasons listed above: the dataset is of insufficient quality and arbitrary assumptions have been made that make it not comparable to other datasets and not suitable for statistical analyses. It is troubling that Love and Ahiaga-Dagbui base their paper on this type of data, resulting in misleading conclusions.

Love and Ahiaga-Dagbui criticize David Uren, the Economics Editor of *The Australian*, for citing our results, and for assuming that the results derive from megaprojects, i.e., projects larger than $1 billion. But if you read Uren, you will see that when he quotes our numbers he explicitly refers to "major rail projects" and not megaprojects, which is a correct reference that includes projects smaller than megaprojects. To criticize Uren on this point is therefore again both wrong and unfair. Furthermore, Love and Ahiaga-Dagbui criticize Uren for saying that "projects with greater cost overruns deliver a benefit shortfall" (p. 359), which Love and Ahiaga-Dagbui believe (without proof) to be an untrue statement. But Uren says no such thing – true or not – and neither do we. Uren correctly quotes Flyvbjerg (2009: 344) for saying, "The projects that are made to look best on paper are the projects that amass the highest cost overruns and benefit shortfalls in reality." Nothing in this statement implies a mechanism according to which projects with greater cost overruns would deliver a benefit shortfall, as Love and Ahiaga-Dagbui imply. The critique of Uren is therefore misguided.

In sum, Love and Ahiaga-Dagbui present no valid and reliable data to counter our finding that 86 percent of large transportation infrastructure investments have cost overrun. Their postulated Myth No. 3 (that our finding is a myth) therefore falls and our finding stands, until such a time that someone presents data to falsify it (see Table 1). Love and Ahiaga-Dagbui also do not present data to counter our finding that average cost overrun in large transportation infrastructure projects is 28 percent in real terms. In fact, the numbers Love and Ahiaga-Dagbui present support this finding, or perhaps even a number somewhat higher than this.

[23] This is why, when accuracy really counts, many consultancies use our data instead of their own.

5. What are the root causes of cost overrun?

Your biggest risk is you, according to behavioral science. The root cause of cost overrun is human bias, psychological and political. Scope changes, complexity, geology, archaeology, bad weather, business cycles, etc. are causes, but not root causes. If you don't solve the problem of cost overrun at the root, you will not end overrun.

Recent developments in behavioral science are causing Kuhnian paradigm shifts in many fields, including project management and forecasting. Love and Ahiaga-Dagbui are on the wrong side of this shift. Incredibly, they ignore 30–40 years of research in behavioral science, including the Nobel-Prize-winning work of Amos Tversky and Daniel Kahneman on heuristics and biases (Tversky and Kahneman, 1974; Gilovich et al., 2002; Kahneman, 2011). There is not one reference or mention of this work in Love and Ahiaga-Dagbui's paper. As long as cost estimation for large transportation infrastructure projects does not take into account the revolutionary results of behavioral science – i.e., as long as cost estimation is understood and practiced in the manner described by Love and Ahiaga-Dagbui – planners and managers will keep getting cost wrong. This is the most fundamental problem with Love and Ahiaga-Dagbui's approach.

Instead of accepting the Kuhnian revolution of behavioral science, Love and Ahiaga-Dagbui propose what they call an "evolutionist" approach to understanding cost overrun, which emphasizes "changes in scope" and "complexity in complex decisions [sic]" as main causes of overrun (pp. 358, 365). Behavioral scientists would agree that scope changes and complexity are relevant to understanding what goes on in capital investment projects, but would not see them as root causes of cost overrun. The root cause of cost overrun, according to behavioral science, is the well-documented fact that planners and managers keep underestimating scope changes and complexity in project after project.

From the point of view of behavioral science, the mechanisms of scope changes, complex interfaces, archaeology, geology, bad weather, business cycles, etc. are not unknown to planners of capital projects, just as it is not unknown to planners that such mechanisms may be mitigated, for instance by reference class forecasting (see below). However, planners often underestimate these mechanisms and mitigation measures, due to overconfidence bias, the planning fallacy, and strategic misrepresentation. In behavioral terms, scope changes etc. are manifestations of such underestimation on the part of planners, and it is in this sense that bias and underestimation are the root causes of cost overrun. But because scope changes etc. are more visible than the underlying root causes, they are often mistaken for the cause of cost overrun. In behavioral terms, the causal chain starts with human bias which leads to underestimation of scope during planning which leads to unaccounted for scope changes during delivery which lead to cost overrun. Scope changes are an intermediate stage in this causal chain through which the root causes manifest themselves. With behavioral science we say to planners, "*Your biggest risk is you.*" It is not scope changes, complexity, etc. in themselves that are the main problem; it is how human beings misconceive and underestimate these phenomena, through overconfidence bias, the planning fallacy, etc. This is a profound and proven insight that behavioral science brings to capital investment planning.

Behavioral science entails a change of perspective: *The problem with cost overrun is not error but bias*, and as long as you try to solve the problem as something it is not (error), you will not solve it. Estimates and decisions need to be de-biased, which is fundamentally different from eliminating error (Kahneman et al., 2011; Flyvbjerg, 2008, 2013a). Furthermore, *the problem is not even cost overrun, it is cost underestimation*. Overrun is a consequence of underestimation, with the latter happening upstream from overrun, often years before overruns manifest. Again, if you try to solve the problem as something it is not (cost overrun), you will fail. You need to solve the problem of cost underestimation to solve the problem of cost overrun. Until you understand these basic insights from behavioral science, you're unlikely to get capital investments right, including cost estimates. Love and Ahiaga-Dagbui clearly do not understand this.

In particular, Love and Ahiaga-Dagbui criticize our use of the terms delusion and deception (Flyvbjerg et al., 2009; Cantarelli et al., 2010a), fools and liars (Flyvbjerg, 2013a), and error and lying (Flyvbjerg et al., 2002) to describe forecasting and forecasters. They say, for example, that our use of the terms fools and liars is "intentionally crafted to be provocative and controversial, has no scientific merit and has been simply contrived to gain attention" (p. 358). Love and Ahiaga-Dagbui have no way, of course, to know what our intentions are. But worse, they fail to mention that the terms fool and liar are not ours but those of Nassim Nicholas Taleb, although we clearly state this in the work Love and Ahiaga-Dagbui cite, by including the following verbatim quote from Taleb (2007: 163, emphasis added):

> "People … who forecast simply because 'that's my job,' knowing pretty well that their forecast is ineffectual, are not what I would call ethical. What they do is no different from repeating lies simply because 'it's my job.' Anyone who causes harm by forecasting should be treated as either *a fool or a liar*. Some forecasters cause more damage to society than criminals."[24]

Love and Ahiaga-Dagbui further fail to mention that we use the terms "delusion" and "deception," etc. in an academically precise manner referring to well-established bodies of theory, with people suffering from delusion explicitly defined as those "who are subject

[24] Recently, for the first time, transportation forecasters have been treated as criminals in a number of lawsuits in Australia and the USA over misleading forecasts (Dezember and Glazer, 2013; Evans, 2010; Hals, 2013; Miller, 2013; Rubin, 2017; Saulwick, 2014; Singh, 2017; Worthington, 2012; Wright, 2014). This has sent shock waves through the international forecasting community.

to 'optimism bias' and the 'planning fallacy'," whereas people operating by means of deception are defined as "those practicing 'strategic misrepresentation,' 'agency,' and the 'conspiracy of optimism'" (Flyvbjerg, 2013a: 772).[25] We see these terms as analytically sharp and as highly communicative concepts that help us understand and reframe what is going on in forecasting, and how to improve its outcomes. That is why we use the terms. We agree with Love and Ahiaga-Dagbui that more explanations of cost overrun exist than delusion and deception, and we discuss these in Cantarelli et al. (2010a). We maintain, however, that explanations are not born equal and that root causes are more important than causes in understanding and curbing overrun.

Love and Ahiaga-Dagbui further claim, again falsely, that, "No evidence at all supports the causal claims of delusion and deception" (p. 366). Love and Ahiaga-Dagbui here conveniently disregard a body of scholarship that runs from Wachs (1989, 1990) over Kain (1990) and Pickrell (1990) to Flyvbjerg et al. (2004), Flyvbjerg et al. (2009), and Cantarelli et al. (2010a), which demonstrates that deception contributes to cost underestimation. For an example in their own backyard, we suggest Love and Ahiaga-Dagbui study the history of the Sydney Opera House, which was built on lies causing a cost overrun of 1400 percent, as we show in Flyvbjerg et al. (2009) and Flyvbjerg (2005). There is a reason why Martin Wachs, one of the most respected and trusted scholars in the field, called his classic paper on the topic "When Planners *Lie* with Numbers" (Wachs, 1989, emphasis added).[26] Unlike Love and Ahiaga-Dagbui, Wachs, Flyvbjerg, Glenting, Rønnest, and more have made the effort to actually interview forecasters about delusion and deception. Wachs reported that in case after case, forecasters told him they had had to "revise" their forecasts many times because they failed to satisfy their superiors. The forecasts had to be "cooked" in order to produce numbers that were suitable to getting projects started. The following is a typical example from Wachs's (1990: 144–145) interviews:

> "a planner admitted to me that he had reluctantly but repeatedly adjusted the patronage figures upward, and the cost figures downward to satisfy a local elected official who wanted to compete successfully for a federal grant. Ironically, and to the chagrin of that planner, when the project was later built, and the patronage proved lower and the costs higher than the published estimates, the same local politician was asked by the press to explain the outcome. The official's response was to say, 'It's not my fault; I had to rely on the forecasts made by our staff, and they seem to have made a big mistake here'."

Wachs (1986: 28, 1990: 146) talks of "nearly universal abuse" of forecasting in this manner. Flyvbjerg et al. (2004: 44) were able to replicate Wachs's results more than a decade later with different data, further confirming the theories of deception. The following planner is typical of how respondents explained the basic mechanism of cost underestimation:

> "You will often as a planner know the real costs. You know that the budget is too low but it is difficult to pass such a message to the counsellors [politicians] and the private actors. They know that high costs reduce the chances of national funding."

Experienced professionals like the interviewed planner know that outturn costs will be higher than estimated costs due to scope changes, complex interfaces, archaeology, geology, bad weather, business cycles, etc., but because of political pressure to secure funding for projects they hold back this knowledge, which is seen as detrimental to the objective of obtaining funding.

Wachs (2013: 112), who pioneered research on deception in transportation infrastructure forecasting, recently summed up more than 25 years of scholarship in the following manner, in stark contrast to Love and Ahiaga-Dagbui:

> "While some scholars believe this [i.e., misleading forecasts] is a simple technical matter involving the tools and techniques of cost estimation and patronage forecasting, there is growing evidence that the gaps between forecasts and outcomes are the results of deliberate misrepresentation and thus amount to a collective failure of professional ethics."

Love and Ahiaga-Dagbui may dislike research results like these, because they identify members of their profession as unethical. It is deeply troubling, however, that people who call themselves scholars would choose to simply ignore the research and postulate a different reality from that documented by it.[27] It reveals a selective approach to evidence that is similar to data picking, i.e., the selection of information to falsely obtain the conclusions one wants.

We further observe that if the inaccuracy of cost estimates were simply a matter of incomplete information and honest errors regarding scope and complexity, as Love and Ahiaga-Dagbui maintain, then one would expect cost inaccuracies to be random or close to random, with overestimates occurring about as frequently as underestimates. We explicitly tested this thesis and falsified it at an overwhelmingly high level of statistical significance ($p < 0.001$) (Flyvbjerg et al., 2002: 286–287). Instead, cost inaccuracies have a striking systemic bias, with underestimates being significantly more common than overestimates, as demonstrated by our studies and those of others as we saw above. Love and Ahiaga-Dagbui provide no data to support their claim that cost estimates suffer from error and not bias. They are even contradicted on this point by their own studies (Love et al., 2012). In the face of such self-contradictory evidence – and confronted with well-documented phenomena like the planning fallacy and overconfidence bias – to postulate randomness in cost underestimation, like Love and Ahiaga-Dagbui do, is misrepresentation, and denial, in the extreme.

Love and Ahiaga-Dagbui argue that our talk of delusion and deception, error or lie, and fools and liars in forecasting, "presents the reader with a false dichotomy; an either/or choice that is practically invalid," in their words (p. 365). Again this is a misreading on

[25] Love and Ahiaga-Dagbui question that there is a rational motive for actors, including forecasters, to lie in project planning and management. In doing so they ignore a strong body of work in economics and management, namely principal-agent theory, which has established exactly this: that lying is often rational in decision making and creates moral hazard.

[26] Full disclosure: Martin Wachs was Bent Flyvbjerg's supervisor when he was a doctoral student at the University of California, Los Angeles. Flyvbjerg also gave the first Annual Martin Wachs Distinguished Lecture at UC Berkeley in 2006.

[27] For other examples of such denial, see Flyvbjerg (2013b, 2015).

the part of Love and Ahiaga-Dagbui. We never claimed that delusion and deception and error and lie cannot occur simultaneously in the same project, so there is no dichotomy. In fact, Flyvbjerg et al. (2009: 180) explicitly emphasize a both/and, in saying, "Delusion and deception are complementary rather than alternative explanations."

Finally, Love and Ahiaga-Dagbui offer the influence of procurement methods as an alternative to our explanations of cost overrun. They write that, "Flyvbjerg et al.… have not considered or acknowledged the influence of procurement methods and other project practices that can have on [sic] large-scale transport infrastructures [sic] projects outturn costs" (p. 364). This is factually incorrect. Flyvbjerg et al. (2004) and Cantarelli and Flyvbjerg (2015) compared procurement in public-private partnership projects, projects under state-owned enterprises, and conventional public sector projects in a study that produced statistically significant results regarding procurement types. In comparison, Love and Ahiaga-Dagbui arbitrarily present numbers without statistical analysis and scientific validity. We agree with Love and Ahiaga-Dagbui that the influence of procurement methods on cost overrun is an interesting area for further research and we have several more such studies in the pipeline. However, (a) the studies must be substantially more rigorous than what Love and Ahiaga-Dagbui present and (b) we see no indication in existing results that choice of procurement method may be considered a root cause of cost overrun, but only a cause, as per our distinction above.

In sum, Love and Ahiaga-Dagbui designate as myth our explanation of cost overrun in terms of psychological and political bias. Instead, they explain cost overrun in terms of errors caused by scope changes and complexity. We have on our side (a) Nobel-Prize-winning theory on heuristics and biases that show planners will systematically underestimate cost resulting in cost overrun, and (b) principal-agent theory, which shows that lying about cost can be rational and is often incentivized in decision making. Even better, we have high-quality data that fit the two sets of theory at an overwhelmingly high level of statistical significance ($p < 0.001$), and we have planners stating on the record that they deliberately underestimated costs to have projects approved and funded, establishing the causal link between deception and cost underestimation that Love and Ahiaga-Dagbui say does not exist. In comparison, Love and Ahiaga-Dagbui offer their untested "evolutionist" theory that (a) does not fit existing data – ours or that of others – because according to the theory inaccuracies in cost estimates would be more or less random, but in reality they are systemically biased, and (b) has resulted in unsuccessful mitigation and containment strategies, according to Love and Ahiaga-Dagbui themselves (see below).

As the evidence stands, Love and Ahiaga-Dagbui's Myth No. 4 (that explanations in terms of bias are a myth) falls and our explanation stands, because it is better supported both theoretically and empirically (see Table 1). If you want to get cost estimates right, you need to unlearn key parts of conventional cost estimation, with its century-long record of getting estimates wrong. Instead you need to learn behavioral science, because behavior is the main problem, not artefacts.

6. How is cost overrun best avoided?

Cost overrun is best avoided by (a) Getting the front-end of capital investments right, including using reference class forecasting or similar methods to establish reliable, de-biased estimates of cost that fit the client's risk appetite, (b) Establishing an incentive structure that encourages all involved to stay on budget, and (c) Hiring a delivery team with a proven track record for the specific type of capital investment in question.

Love and Ahiaga-Dagbui admit that the conventional strategies to curb cost overrun have failed, when they write "While considerable inroads have been made by the evolutionists to explain cost overrun … the mitigation and containment strategies that have been developed to combat this phenomenon have fallen short of their intended goal" (p. 358). We agree, but given this admission of failure for conventional strategies, we find it difficult to understand Love and Ahiaga-Dagbui's animosity toward alternatives, including reference class forecasting, which has been documented to produce better outcomes.

Data accumulated over the past two to three decades show with overwhelming statistical significance that it is not the fact that budgets are wrong that needs explaining, but the fact that the vast majority of budgets are wrong in the same direction, namely underestimation. Our data go back 86 years and for that period the bias in cost forecasting has been constant. Forecasters are "predictably irrational" in the words of Ariely (2010). Taking this known and highly predictable bias into account in forecasting is the single most important thing planners and managers can do to de-bias forecasts and make them more accurate, as pointed out by Kahneman (2011), Lovallo and Kahneman (2003), and Kahneman and Tversky (1979). Several governments and corporations are already doing this, including in the USA, China, Britain, Scandinavia, and Switzerland. However, as long as people like Love and Ahiaga-Dagbui keep conceiving of the forecasting problem in terms of error instead of bias they block this insight and the improvement in forecasting accuracy it would bring.

Nowhere is Love and Ahiaga-Dagbui's misconstrual of the forecasting problematic – and their limited understanding of statistics and behavioral science – expressed more clearly than in their comments on reference class forecasting (RCF), a method originally proposed by Kahneman and Tversky (1979) to de-bias predictions along the lines described above. Love and Ahiaga-Dagbui write about RCF, "To simply assume that a given project is comparable to past and completed projects and that a lump sum up-lift could be added to account for all uncertainties is a gross oversimplification of reality" (p. 366). We agree it is a simplification, like any forecast will be. But Love and Ahiaga-Dagbui overlook the fact that this simplification has been documented to produce better estimates on average than any other simplification (Kahneman and Lovallo, 1993; Batselier and Vanhoucke, 2016; Chang et al., 2016; Awojobi and Jenkins, 2016). "All models are wrong, but some are useful," as famously said by Box (1979: 202). We see RCF as useful, because it results in better forecasting accuracy than the models of conventional cost estimation, which we see as wrong, because they produce not only error but systemic bias.

In contrast, Love and Ahiaga-Dagbui claim that by coming up with a less simplified and more detailed account of costs they can better the results of RCF. "[A]s information becomes available the reliability of an estimate improves," they state as a general principle, equating more information with more detail (p. 362). Love and Ahiaga-Dagbui provide no proof of their claim, which is

understandable because proof does not exist – quite the opposite. It may sound intuitively right that more detail would lead to higher accuracy, but the claim ignores Occam's razor and formal work on statistical model selection (Akaike, 1970, 1998; Box et al., 2015; Lee, 1973; Krugman, 2000; Smith, 1997; Stoica and Söderström, 1982; Tukey, 1961) and human decision making (Czerlinski et al., 1999; Gigerenzer and Gaissmaier, 2011; Marewski et al., 2010), which demonstrates that when two models fit the data the simpler one is generally more accurate. Box (1976: 792) elegantly sums up the situation: "Just as the ability to devise simple but evocative models is the signature of the great scientist so overelaboration and overparameterization is often the mark of mediocrity."

Love and Ahiaga-Dagbui's claim is similar to believing you can beat the basic odds of a Las Vegas casino by a more detailed understanding of how gambling works. You cannot. By taking an outside view, rather than a more detailed inside view that invites optimism, RCF gives you the basic odds of cost estimation, and on average you will be better off sticking to these odds than anything else you can think up. This is plausibly argued by behavioral science and empirically documented by Batselier (2016), Batselier and Vanhoucke (2017), Bordley (2014), Kim and Reinschmidt (2011), Liu and Napier (2010), and Liu et al. (2010).

Love and Ahiaga-Dagbui further claim that, "The production of an estimate demands knowledge of what will occur … The acquisition of such knowledge is dependent upon the completeness of the information made available" (p. 363). Again this claim is wrong and it conveys a deterministic type of thinking we would have thought extinct in academia after the probabilistic revolution has shown that nothing is deterministic in nature. The claim that we must know "what will occur" to make an estimate is akin to insisting on understanding the world in terms of Newtonian physics after quantum mechanics, something you would not get away with in physics, but something Love and Ahiaga-Dagbui apparently think still goes in their field. You do not need to know what *will* occur, nor do you need "completeness of … information," to make high-quality forecasts. You need to know the odds of what *may* occur, and RCF provides those odds. As long as Love and Ahiaga-Dagbui insist on producing their more detailed forecasts, based on conventional cost estimation, they will continue to produce inferior forecasts.[28]

We understand that our line of argument may be unsettling for Love and Ahiaga-Dagbui, because it shows that much of what they do can be bettered by simpler and theoretically sounder methods. But such is innovation in cost estimation today. Accept it and you may prosper; deny it and you will be left irrelevant. However, conventional cost estimation and RCF need not be at loggerheads. We have developed an approach in which they supplement each other, with RCF correcting the biases of the conventional approach, which contributes a useful breakdown of project components needed for effective delivery. This supplementary approach has been applied in practice on dozens, if not hundreds, of projects and is described in Flyvbjerg et al. (2004), Flyvbjerg (2008), Flyvbjerg et al. (2009), Flyvbjerg et al. (2004), Batselier (2016), and Batselier and Vanhoucke (2017). We encourage readers, and especially conventional cost engineers, to try out this combined approach on their next projects and see the difference it makes to accuracy.

Love and Ahiaga-Dagbui claim that "*Reference Class Forecasting* … utilizes a Normal distribution" (p. 366, emphasis and caps in original). This is factually wrong. RCF uses the empirical distribution in the reference class, whatever it is. For large infrastructure projects, typically the distribution is not normal but asymmetrical and fat-tailed.

In addition, Love and Ahiaga-Dagbui write that "it is more appropriate to use the median rather than the mean, which Flyvbjerg (2008) utilizes when applying RCF" (p. 366). This is factually wrong again. First, the median and the mean may both be used by forecasters, depending on what level of certainty they wish to achieve for their forecast, as explained in Flyvbjerg (2008). The median, also called the P50, is used by forecasters who are willing to accept a fifty-fifty risk of cost overrun. The mean is used by portfolio managers, because in this case projects in the portfolio that go over budget will be balanced by projects that go under. For distributions with a fat upper tail, like cost overrun, the mean will be significantly higher than the median, so using one instead of the other will significantly influence the risks accepted by the forecast. Specifically, using the median, as Love and Ahiaga-Dagbui recommend, would significantly underestimate risk, incurring risks higher than those accepted by portfolio managers. Second, when clients manage just one or a few projects, which is typical for megaprojects, they are often more risk averse than portfolio managers. In this case neither median nor mean is used in RCF, but higher P-values, often the P80, which indicates 80 percent certainty of staying within budget and 20 percent risk of going over (Flyvbjerg, 2008: 13). In sum, we do not use the mean in RCF, we use the value that corresponds to clients' risk appetite, and the client decides. For the very large projects we typically work with, this mostly corresponds to the more conservative P80-value, but if clients are even more conservative than this, a higher P-value is used. So far the highest P-value we have used in practical RCF is the P95.

Love and Ahiaga-Dagbui further write that "an estimate for a large infrastructure project should include the estimated range of uncertainty," and they seem to think this is best measured by the standard deviation (pp. 365–66). Again this is wrong and Love and Ahiaga-Dagbui here contradict themselves. Above we saw they argue that the mean is not a good summary measure for risk distributions. But standard deviations are linked to the mean, so to be against means but for standard deviations as measures of uncertainty is inconsistent and constitutes yet another statistical error on the part of Love and Ahiaga-Dagbui. More fundamentally, distributions of cost overrun for large infrastructure projects are asymmetrical and fat-tailed, as said. For such distributions the standard deviation is not a good measure of uncertainty. The standard deviation underrepresents fat tails and gives the impression that distributions are symmetric, i.e., that overruns and underruns around the central value are equally likely, which is not the case for large infrastructure projects. Kahneman and Tversky (1979) advocate instead presentation of *the full distributional information* as the preferred and most transparent option, which is what we do when we do RCFs (Flyvbjerg et al., 2004; Flyvbjerg et al., 2016).

[28] Love and Ahiaga-Dagbui also claim that, "as design process [sic] become [sic] digitized, enabled by Building Information Modelling [BIM], cost estimates will improve" (p. 363). We say, do not hold your breath. BIM has been around for 25 years, and although examples exist of effective use of BIM results have in general been disappointing. It will take fundamental disruption beyond BIM to bring the construction industry into the Digital Age, in our judgment.

Aimed at people like Love and Ahiaga-Dagbui, Taleb (2014: 535) observes that the standard deviation "does more harm than good – particularly with the growing class of people … mechanistically applying statistical tools to scientific problems." Taleb recommends the use of *the median absolute deviation* as a more robust measure of variability. In any case, no one familiar with the actual distribution of cost overrun in large infrastructure projects would recommend the standard deviation as a measure of uncertainty. The fact that Love and Ahiaga-Dagbui, and similar-minded forecasters, advocate and use this measure helps explain why they keep getting cost estimates wrong.

In their perhaps most blatant demonstration of statistical ineptitude, Love and Ahiaga-Dagbui talk about "fitting … data to the distribution" (p. 365). Somebody who knows statistics would never write like this. From a statistician's point of view data are immutable and can therefore not be meddled with in the manner suggested by Love and Ahiaga-Dagbui. Distributions are fitted to the data, or a mathematical transformation of them, not the other way around.[29]

As a final point regarding RCF, Love and Ahiaga-Dagbui cite the Edinburgh Tram as "[a]n example where RCF was applied and an inappropriate distribution used" (p. 366). This is wrong. First, the cost overrun distribution used for the first Edinburgh Tram RCF reflects the historical data for all comparable projects for which data were available at the time of doing this RCF, as the theory of RCF says it must.[30] Second, Love and Ahiaga-Dagbui indicate that the Edinburgh Tram RCF was in error because actual cost was higher than that forecasted by the RCF. This is a naive interpretation of RCF, which again reveals Love and Ahiaga-Dagbui's deterministic bent and lack of understanding of uncertainty. Even a P80 forecast (80 percent chance of staying within budget), like that initially used for the Edinburgh Tram, has a risk of overrun, namely 20 percent. This means that one in five times the forecast will be exceeded, which, then, should come as no surprise when it happens and does not invalidate the approach of estimation, as Love and Ahiaga-Dagbui indicate. Third, although the Edinburgh Tram planners did initially adopt an RCF approach they later turned away from this and back to a fully conventional approach imbued with optimism and inaccuracy (Flyvbjerg and Budzier, 2018). Finally, Love and Ahiaga-Dagbui confuse the numbers they cite for the Edinburgh Tram by comparing monetary values that (a) are not given in the same year's prices, and (b) cover different cost items, undermining their conclusions. It is unsettling to see supposedly experienced cost engineers make such basic mistakes of comparing apples and oranges. For up-to-date and consistent numbers on the Edinburgh Tram, see Flyvbjerg and Budzier (2018).[31] In sum, if we ever saw an example of optimism bias and a project that would have benefitted from a consistent RCF, the Edinburgh Tram is it. You do not have to take our word for it. Read the Edinburgh Tram Inquiry.[32]

Love and Ahiaga-Dagbui rightly say evidence should decide truth claims. Today, a dozen independent evaluations exist with evidence that supports the accuracy of RCF over other estimation methods, for large and small projects alike (Chang et al., 2016; Awojobi and Jenkins, 2016; Batselier, 2016; Batselier and Vanhoucke, 2017; Bordley, 2014; Kim and Reinschmidt, 2011; Liu and Napier, 2010; Liu et al., 2010). Here is the conclusion from one such evaluation, covering construction projects:

> "The conducted evaluation is entirely based on real-life project data and shows that RCF indeed performs best, for both cost and time forecasting, and therefore supports the practical relevance of the technique" (Batselier and Vanhoucke, 2016: 36).

7. Conclusions: good and bad practice

Above we refuted, one by one, Love and Ahiaga-Dagbui's four myths about cost overrun, summarized in Table 1. Table 3 shows good and bad practice in understanding and curbing cost overrun in large capital investment projects. We argued above that Love and Ahiaga-Dagbui's paper epitomizes bad practice. We contrasted this with good practice, which we spelled out in terms of how to best understand and curb cost overrun. Good practice entails:

1. *Consistent definition and measurement of overrun*, as actual cost in percent (or ratio) of estimated cost, with cost measured in the local currency, constant prices, and against a congruent baseline; in contrast to Love and Ahiaga-Dagbui who inconsistently mix and compare studies with different baselines and price levels.
2. *Data collection that includes all observations of overrun* that are valid, reliable, and comparable; as opposed to Love and Ahiaga-Dagbui who include in their paper non-valid data from consultancies, idiosyncratically sampled data that do not compare to other data, data with arbitrarily removed outliers, data that mix population and sample, data from populations that do not compare, data from small and big projects, and data based on different baselines, rendering their arguments invalid.
3. *Acknowledgment that cost overrun is fat-tailed with a significant overincidence of overrun to underrun, indicating substantial upper tail risk*; in contrast to Love and Ahiaga-Dagbui who ignore or underplay the systemic bias in cost overrun and talk of overrun in terms of error and randomness, against all evidence, consequently underestimating cost risk.

[29] Here Love and Ahiaga-Dagbui also claim, "The Flyvbjerg et al. (2002) study simply relies on a Normal distribution information [sic] in their dataset and measures of *p*-values to reach the sweeping conclusions made" (p. 365). Again this is wrong. Anyone caring to read the study would see we use non-parametric methods where normality cannot be assumed. Moreover, if Love and Ahiaga-Dagbui understood the Central Limit Theorem, they would know that (a) even if individual measurements (here cost overrun) do not come from a normal distribution, the average of many measurements will tend to follow normality with good approximation, and (b) this justifies using normality in many statistical tests.

[30] The reference class and cost overrun distribution used for the first Edinburgh Tram RCF was developed by Flyvbjerg, Glenting, and Rønnest in collaboration with the UK Department for Transport.

[31] Flyvbjerg and Budzier (2018) was written as part of Flyvbjerg serving as expert witness for the Edinburgh Tram Inquiry.

[32] Proceedings from the Edinburgh Tram Inquiry may be found at http://www.edinburghtraminquiry.org.

Table 3
Good and bad practice for understanding and curbing cost overrun.

	Good Practice	Bad Practice
1. Definition and measurement	Cost overrun is defined as actual cost in percent (or ratio) of estimated cost, with cost measured in the local currency, constant prices, and against a consistent baseline. The baseline must reflect what you want to measure, e.g., whether (a) the decision to build was well informed, or (b) contractors deliver to their budget.	Treating overruns as comparable that are not because overrun was inconsistently defined and measured, e.g., mix of baselines, price levels, or currencies, which is common. Inconsistencies regarding baselines etc. make it misleading to compare overrun across project types, organizations, geographies, historical periods, etc.
2. Data	All comparable observations of overrun for which valid and reliable data are available should be included, so that no distributional information goes to waste. Observations may constitute the whole population, or they may be from a sample, which should be large enough to demonstrate statistical significance. Possible biases arising from sampling must be carefully assessed, including their impact on results.	"Garbage in, garbage out," as a result of: • Including idiosyncratically sampled data • Excluding outliers • Mixing data based on different baselines • Including low-quality data from consultancies • Confusing population and sample • Mixing data from fundamentally different populations • Assuming larger populations require larger samples • Thinking meta-analysis constitutes "piggy-backing" • Assuming data from secondary sources are less valuable than primary data
3. Size and frequency	High-quality studies baselined at the decision to build show: • The vast majority of projects have overrun • Average and median overrun are positive and significantly different from zero, with substantial variation between investment types • The distribution of overruns is fat-tailed to the right	• Basing results on bad data (see above) • Falsely assuming error and randomness for results that are systemically biased • Disregarding fat tails • Denying or dismissing results because you don't like them • Assuming comparability of results that are not comparable
4. Root causes	Your biggest risk is you, says behavioral science. Cost overrun is not caused by error, but by bias, psychological and political. As long as you try to solve the problem of overrun as an error, you will not succeed. Cost overrun is not even the problem. Cost underestimation is.	Endlessly repeating the fallacy that cost overrun is an error caused by scope changes, complexity, geology, etc. Disregarding, deliberately or not, scholarship in behavioral science, including on optimism and moral hazard.
5. Solutions	De-biasing estimates of cost by using reference class forecasting or similar methods that build on behavioral science.	Endlessly producing inaccurate and systemically biased forecasts using conventional cost estimation.

4. *Acknowledgment that the root cause of cost overrun is human bias, which leads to underestimation of scope during planning which leads to scope changes during delivery which lead to cost overrun*; as opposed to Love and Ahiaga-Dagbui who dismiss the findings of behavioral science as "fake news" and explain overrun in technical terms, directly caused by scope changes and complexity.
5. *De-biasing estimates of cost by using reference class forecasting or similar methods that build on behavioral science*; in contrast to Love and Ahiaga-Dagbui who use conventional methods with a century-long track record of inaccuracy and systemic bias. But conventional methods and RCF need not be at loggerheads. RCF may be used to correct the biases of the conventional approach, which in turn contributes a useful breakdown of project components needed for effective delivery. We recommend this combined approach, which has a documented track record of high accuracy.

In addition to being exemplars of bad practice for understanding and curbing cost overrun, Love and Ahiaga-Dagbui misrepresent our work in their futile attempts at discrediting behavioral explanations of overrun. We have countered the worst misrepresentations above. Alone and taken together the misrepresentations constitute a fabricated version of our findings, either (a) by presenting direct falsehoods, as when Love and Ahiaga-Dagbui claim we cherry-pick data for our studies when in fact we include all valid and reliable data, making data-picking impossible, or (b) through omissions, as when they postulate that no evidence has demonstrated that deception contributes to cost underestimation when in fact Wachs (1989, 1990), Flyvbjerg et al. (2004), and Flyvbjerg (2005) document such evidence. The fabricated version of our work has little semblance with the actual work and in several cases stands in direct opposition to it, for instance, when Love and Ahiaga-Dagbui claim that reference class forecasting utilizes a normal distribution, when in fact it uses the empirical distribution of the reference class of projects, which is typically fat-tailed.

Nevertheless, we appreciate Love and Ahiaga-Dagbui's acknowledgment of the impact of our work and the opportunity to clarify what good and bad practice is regarding cost overrun. We understand their aversion to the behavioral revolution. After all, it leaves obsolete key parts of what they know and do. That is normal for Kuhnian paradigm shifts. We encourage Love and Ahiaga-Dagbui to get on the right side of the behavioral shift and face up to what the data show: cost overrun is systemic, not random; bias is the problem, not error; behavior is the issue, not artefacts.

References

Akaike, Hirotogu, 1970. On a semi-automatic power spectrum estimation procedure. In: Proceedings of the 3rd Hawaii International Conference on System Sciences, pp. 974–977.

Akaike, Hirotogu, 1998. Information theory and an extension of the maximum likelihood principle. In: Selected Papers of Hirotugu Akaike, Springer, New York, pp. 199-213.

Ansar, Atif, Flyvbjerg, Bent, Budzier, Alexander, Lunn, Daniel, 2014. Should we build more large dams? The actual costs of hydropower megaproject development. Energy Policy 69, 43–56.

Ansar, Atif, Flyvbjerg, Bent, Budzier, Alexander, Lunn, Daniel, 2016. Does infrastructure investment lead to economic growth or economic fragility? Evidence from China. Oxford Rev. Econ. Policy 32 (3, autumn), 360–390.

Ariely, Dan, 2010. Predictably Irrational: The Hidden Forces That Shape Our Decisions. HarperCollins, New York.

Awojobi, Omotola, Jenkins, Glenn P., 2016. Managing the cost overrun risks of hydroelectric dams: an application of reference class forecasting techniques. Renew. Sustain. Energy Rev. 63, 19–32.

Batselier, Jordy, 2016. Empirical Evaluation of Existing and Novel Approaches for Project Forecasting and Control. Doctoral dissertation. University of Ghent, Ghent.

Batselier, Jordy, Vanhoucke, Mario, 2016. Practical application and empirical evaluation of reference class forecasting for project management. Project Manage. J. 47 (5), 36–51.

Batselier, Jordy, Vanhoucke, Mario, 2017. Improving project forecast accuracy by integrating earned value management with exponential smoothing and reference class forecasting. Int. J. Project Manage. 35 (1), 8–43.

Bordat, Claire, McCullouch, Bob G., Sinha, Kumares C., Labi, Samuel, 2004. An Analysis of Cost Overruns and Time Delays of INDOT Projects. Purdue University, Joint Transportation Research Program, West Lafayette, IN.

Bordley, Robert F., 2014. Reference class forecasting: resolving its challenge to statistical modeling. Am. Stat. 68 (4), 221–229.

Box, George E.P., 1976. Science and statistics. J. Am. Stat. Assoc. 71 (356), 791-799.

Box, George E.P., 1979. Robustness in the strategy of scientific model building. In: Launer, Robert L., Wilkinson, Graham N. (Eds.), Robustness in Statistics. Academic Press, Cambridge, MA, pp. 201–236.

Box, George E., Jenkins, Gwilym M., Reinsel, Gregory C., Ljung, Greta M., 2015. Time Series Analysis: Forecasting and Control. John Wiley & Sons, New York.

Cantarelli, Chantal C., Flyvbjerg, Bent, Molin, Eric J.E., van Wee, Bert, 2010a. Cost overruns in large-scale transportation infrastructure projects: explanations and their theoretical embeddedness. Eur. J. Transp. Infrastruct. Res. 10 (1), 5–18.

Cantarelli, Chantal C., Flyvbjerg, Bent, van Wee, Bert, Molin, Eric J.E., 2010b. Lock-in and its influence on the project performance of large-scale transportation infrastructure projects: investigating the way in which lock-in can emerge and affect cost overruns. Environ. Plan. B: Plan. Des. 37, 792–807.

Cantarelli, Chantal C., Flyvbjerg, Bent, 2015. Decision making and major transport infrastructure projects: the role of project ownership. In: Hickman, Robin, Bonilla, David, Givoni, Moshe, Banister, David (Eds.), Handbook on Transport and Development. Edward Elgar, Cheltenham, UK and Northampton, MA, August, pp. 380–393.

Cantarelli, Chantal C., Flyvbjerg, Bent, Buhl, Søren L., 2012a. Geographical variation in project cost performance: the Netherlands versus worldwide. J. Transp. Geogr. 24, 324–331.

Cantarelli, Chantal C., Molin, Eric J.E., van Wee, Bert, Flyvbjerg, Bent, 2012b. Characteristics of cost overruns for dutch transport infrastructure projects and the importance of the decision to build and project phases. Transp. Policy 22, 49–56.

Cantarelli, Chantal C., Molin, Eric J.E., van Wee, Bert, Flyvbjerg, Bent, 2012c. Different cost performance: different determinants? The case of cost overruns in Dutch transport infrastructure projects. Transp. Policy 22, 88–95.

Chang, Welton, Chen, Eva, Mellera, Barbara, Tetlock, Philip, 2016. Developing expert political judgment: the impact of training and practice on judgmental accuracy in geopolitical forecasting tournaments. Judg. Decision Mak. 11 (5), 509–526.

Czerlinski, Jean, Gigerenzer, Gerd, Goldstein, Daniel G., 1999. "How Good are Simple Heuristics? In Simple Heuristics that Make Us Smart. Oxford University Press, Oxford, pp. 97–118.

Dantata, Nasiru A., Touran, Ali, Schneck, Donald C., 2006. Trends in US rail transit project cost overrun. In: Transportation Research Board Annual Meeting. National Academies, Washington, DC.

Dezember, Ryan, Glazer, Emily, 2013. Drop in traffic takes toll on investors in private roads. Wall Street J. 20.

Ellis, Ralph, Pyeon, Jae-Ho, Herbsman, Zohar, Minchin, Edward, Molenaar, Keith, 2007. Evaluation of Alternative Contracting Techniques on FDOT Construction Projects. University of Florida, Department of Civil and Coastal Engineering, Gainesville, FLA.

Evans, Anthony G., 2010. Declaration of Anthony G. Evans, Chief Financial Officer of South Bay Expressway, L.P., in Support of the Debtors' Chapter 11 Petitions and First Day Motions," case 10-04516-LA11, filed 22 March 2010, United States Bankruptcy Court Southern District of California.

Federal Transit Administration, 2003. Predicted and Actual Impacts of New Starts Projects: Capital Cost, Operating Cost and Ridership Data, Office of Planning and Environment with support from SG Associates, Inc. Author, Washington, DC.

Federal Transit Administration, 2008. Before and After Studies of New Starts Projects, report to Congress. Author, Washington, DC.

Flyvbjerg, Bent, 2005. Design by deception: the politics of megaproject approval. Harvard Design Magazine, no. 22, Spring/Summer, pp. 50–59.

Flyvbjerg, Bent, 2008. Curbing optimism bias and strategic misrepresentation in planning: reference class forecasting in practice. Eur. Plan. Stud. 16 (1), 3–21.

Flyvbjerg, Bent, 2009. Survival of the unfittest: why the worst infrastructure gets built, and what we can do about it. Oxford Rev. Econ. Policy 25 (3), 344–367.

Flyvbjerg, Bent, 2013a. Quality control and due diligence in project management: getting decisions right by taking the outside view. Int. J. Project Manage. 31 (5), 760–774.

Flyvbjerg, Bent, 2013b. How planners deal with uncomfortable knowledge: the dubious ethics of the American Planning Association. Cities 32(June), 157–163; with comments by Ali Modarres, David Thacher, and Vanessa Watson (June 2013), and Richard Bolan and Bent Flyvbjerg (February 2015).

Flyvbjerg, Bent, 2014a. What you should know about megaprojects and why: an overview. Project Manage. J. 45 (2), 6–19.

Flyvbjerg, Bent, 2015. More on the dark side of planning: response to Richard Bolan. Cities 42, 276–278.

Flyvbjerg, Bent, 2016. The fallacy of beneficial ignorance: a test of Hirschman's Hiding hand. World Dev. 84, 176–189.

Flyvbjerg, Bent, Budzier, Alexander, 2011. Why your IT project may be riskier than you think. Harvard Bus. Rev. September, 601–603.

Flyvbjerg, Bent, Budzier, Alexander, 2018. Report for the Edinburgh Tram Inquiry. The Edinburgh Tram Inquiry, Edinburgh.

Flyvbjerg, Bent, Glenting, Carsten, Rønnest, Arne, 2004. Procedures for Dealing with Optimism Bias in Transport Planning: Guidance Document. UK Department for Transport, London.

Flyvbjerg, Bent, Garbuio, Massimo, Lovallo, Dan, 2009. Delusion and deception in large infrastructure projects: two models for explaining and preventing executive disaster. California Manage. Rev. 51 (2), 170–193.

Flyvbjerg, Bent, Holm, Mette K. Skamris, Buhl, Søren L., 2002. Underestimating costs in public works projects: error or lie? J. Am. Plan. Assoc. 68 (3), 279–295.

Flyvbjerg, Bent, Holm, Mette K. Skamris, Buhl, Søren L., 2004. What causes cost overrun in transport infrastructure projects? Transp. Rev. 24 (1), 3–18.

Flyvbjerg, Bent, Kao, Tsung-Chung, Budzier, Alexander, 2014. Report to the Independent Board Committee on the Express Rail Link Project. In: MTR Independent Board Committee, Second Report by the Independent Board Committee on the Express Rail Link Project (Hong Kong: MTR), pp. A1–A122.

Flyvbjerg, Bent, Stewart, Allison, Budzier, Alexander, 2016. The Oxford Olympics Study 2016: Cost and Cost Overrun at the Games. Saïd Business School Working Papers. University of Oxford, Oxford.

Fouracre, P.R., Allport, R.J., Thomson, J.M., 1990. The Performance and Impact of Rail Mass Transit in Developing Countries, Research Report no. 278. Transport and Road Research Laboratory, Crowthorne.

Gigerenzer, Gerd, Gaissmaier, Wolfgang, 2011. Heuristic decision making. Annu. Rev. Psychol. 62 (1), 451–482.

Gilovich, Thomas, Griffin, Dale, Kahneman, Daniel (Eds.), 2002. Heuristics and Biases: The Psychology of Intuitive Judgment. Cambridge University Press, Cambridge.

Hals, Tom, 2013. Detroit Windsor Tunnel Operator Files for Bankruptcy. Reuters, 25 July 2013.

Hinze, Jimmie, Selstead, Gregory A., 1991. Analysis of WSDOT Construction Cost Overruns. Washington State Department of Transportation, Olympia, WA.
Kahneman, Daniel, 2011. Thinking, Fast and Slow. Farrar, Straus and Giroux, New York.
Kahneman, Daniel, Lovallo, Dan, 1993. Timid choices and bold forecasts: a cognitive perspective on risk taking. Manage. Sci. 39, 17–31.
Kahneman, Daniel, Lovallo, Dan, Sibony, Olivier, 2011. Before you make that big decision. Harvard Bus. Rev. June, 51–60.
Kahneman, Daniel, Tversky, Amos, 1979. Intuitive prediction: biases and corrective procedures. In: Makridakis, S., Wheelwright, S.C. (Eds.), Studies in the Management Sciences: Forecasting, vol. 12. North Holland, Amsterdam, pp. 313–327.
Kain, John F., 1990. Deception in Dallas: strategic misrepresentation in rail transit promotion and evaluation. J. Am. Plan. Assoc. 56 (2), 184–196.
Kim, Byung-Cheol, Reinschmidt, Kenneth F., 2011. Combination of project cost forecasts in earned value management. J. Constr. Eng. Manage. 137 (11), 958–966.
Krugman, Paul, 2000. How complicated does the model have to be? Oxford Rev. Econ. Policy 16 (4), 33–42.
Leavitt, Dan, Ennis, Sean, McGovern, Pat, 1993. The Cost Escalation of Rail Projects: Using Previous Experience to Re-evaluate the CalSpeed Estimates, Working Paper 567. Institute of Urban and Regional Development, University of California, Berkeley.
Lee Jr., , Douglass, B., 1973. Requiem for large-scale models. J. Am. Inst. Plan. 39 (3), 163–178.
Lee, Jin-Kyung, 2008. Cost overrun and cause in Korean social overhead capital projects: roads, rails, airports, and ports. J. Urban Plann. Dev. 134 (2), 59–62.
Liu, Li, Napier, Zigrid, 2010. The accuracy of risk-based cost estimation for water infrastructure projects: preliminary evidence from Australian Projects. Constr. Manage. Econ. 28 (1), 89–100.
Liu, Li, Wehbe, George, Sisovic, Jonathan, 2010. The accuracy of hybrid estimating approaches: a case study of an Australian State road and traffic authority. Eng. Econ. 55, 225–245.
Lovallo, Dan, Clarke, Carmine, Camerer, Colin, 2012. Robust analogizing and the outside view: two empirical tests of case-based decision making. Strateg. Manag. J. 33, 496–512.
Lovallo, Dan, Kahneman, Daniel, 2003. Delusions of success: how optimism undermines executives' decisions. Harvard Bus. Rev. July Issue, 56–63.
Love, Peter E.D., Ahiaga-Dagbui, Dominic D., 2018. Debunking fake news in a post-truth era: the plausible untruths of cost underestimation in transport infrastructure projects. Transp. Res. A: Policy Pract. 113, 357–368.
Love, Peter E.D., Sing, Chun-Pung, Wang, Xiangyu, Irani, Zahir, Thwala, Didibhuku W., 2012. Overruns in transportation infrastructure projects. Struct. Infrastruct. Eng. 10 (2), 141–159.
Marewski, Julian N., Gaissmaier, Wolfgang, Gigerenzer, Gerd, 2010. Good judgments do not require complex cognition. Cogn. Process. 11 (2), 103–121.
Miller, Eric, August 2013. American roads LLC files for bankruptcy. Transp. Top. 5.
National Audit Office, NAO, 1992. Department of Transport: Contracting for Roads. National Audit Office, London.
Nijkamp, Peter, Ubbels, Barry, 1999. How reliable are estimates of infrastructure costs? A comparative analysis. Int. J. Transp. Econ. 26 (1), 23–53.
Pickrell, Don, 1990. Urban rail transit projects: forecast versus actual ridership and cost. US Department of Transportation, Washington, DC.
Riksrevisionsverket, 1994. Infrastrukturinvesteringar: En kostnadsjämförelse mellan plan och utfall i 15 större projekt inom Vägverket och Banverket, RRV 1994:23. Riksrevisionsverket, Stockholm.
Rubin, Debra K., 2017. Suit over Failed Brisbane P3 Focuses on Traffic Projections. Engineering News-Record, 18 October.
Saulwick, Jacob, 2014. Trial to start on $144 million lane cove tunnel debacle. Sydney Morning Herald, 10 August.
Singh, Anil C., 2017. Syncora Guar. Inc. v. Alinda Capital Partners LLC. 2017 NY Slip Op 30288(U), Supreme Court, New York County: Docket Number 651258/2012.
Smith, Stanley K., 1997. Further thoughts on simplicity and complexity in population projection models. Int. J. Forecast. 13, 557–565.
Steele, Mark D., Huber, William, 2004. Exploring data to detect project problems. AACE Int. Trans. 32 (12) 21.1–21.7.
Stoica, Peter, Söderström, Torsten, 1982. On the parsimony principle. Int. J. Control 36 (3), 409–418.
Taleb, Nassim Nicholas, 2007. The Black Swan: The Impact of the Highly Improbable. Penguin, London and New York.
Taleb, Nassim Nicholas, 2014. Standard Deviation. In: Brockman, J. (Ed.), This Idea Must Die: Scientific Theories That Are Blocking Progress. Harper Perennial, New York, pp. 535–537.
Terrill, Marion, Coates, Brendan, Danks, Lucille, 2016. Cost overruns in Australian Transport Infrastructure Projects. In: Proceedings of the Australasian Transport Research Forum, 16–18 November, Melbourne, Australia.
Terrill, Marion, Danks, Lucille, 2016. Cost Overruns in Transport Infrastructure (Melbourne, Australia: Grattan Institute). The title page of the report lists Marion Terrill as sole author of the report; however, page two of the report mentions Lucille Danks as co-author and Brendan Coates, Owain Emslie, Hugh Parsonage, and James Button as having made substantial contributions.
Thurgood, Glen S., Walters, Lawrence C., Williams, Gerald R., Dale Wright, N., 1990. Changing environment for highway construction: the Utah experience with construction cost overruns. Transp. Rec. 1262, 121–130.
Tukey, John W., 1961. Discussion, emphasizing the connection between analysis of variance and spectrum analysis. Technometrics 3 (2), 191–219.
Tversky, Amos, Kahneman, Daniel, 1974. Judgment under uncertainty: heuristics and biases. Science 185 (4157), 1124–1131.
Vidalis, S.M., Najafi, F.T., 2002. Cost and Time Overruns in Highway Construction. University of Florida, Department of Civil and Coastal Engineering, Gainesville, FLA.
Wachs, Martin, 1986. Technique vs. advocacy in forecasting: a study of rail rapid transit. Urban Resour. 4 (1), 23–30.
Wachs, Martin, 1989. When planners lie with numbers. J. Am. Plan. Assoc. 55 (4), 476–479.
Wachs, Martin, 1990. Ethics and advocacy in forecasting for public policy. Bus. Prof. Ethics J. 9 (1 and 2), 141–157.
Wachs, Martin, 2013. The past, present, and future of professional ethics in planning. In: Carmon, Naomi, Fainstein, Susan S. (Eds.), Policy, Planning, and People: Promoting Justice in Urban Development. University of Pennsylvania Press, Philadelphia, PA, pp. 101–119.
Walmsley, D.A., Pickett, M.W., 1992. The Cost and Patronage of Rapid Transit Systems Compared With Forecasts, Research Report 352. Transport Research Laboratory, Crowthorne, UK.
Welde, Morten, Odeck, James, 2017. Cost escalation in the front-end of projects: empirical evidence from Norwegian Road Projects. Transp. Rev. 37 (5), 612–630.
Wright, Robert, 2014. US Infrastructure Crisis Drives Toll Road Bankruptcy. Financial Times, 21 September 2014.
World Bank, 1994. World Development Report 1994: Infrastructure for Development. Oxford University Press, Oxford.
Worthington, Elise, 2012. Clem7 Tunnel Investors Launch $150m Class Action. ABC News, 1 June.