PUBLIC SECTOR TENDERING – ISSUES AND ANALYSIS

Alan S White1

School of Aerospace, Civil & Mechanical Engineering, University of New South Wales at the Australian Defence Force Academy. Northcott Drive, Campbell. ACT. Australia 2600

The absence of a robust tendering theory and a companion body of knowledge, has nurtured a public sector environment which promulgates problematic mathematical processes and mandates procedures which make the implementation of superior methods unlikely. Current policies tend to be insensitive to change except in the face of judicial reasoning. The interplay of auction theory, game theory, decision theory, domain ontology and agreement design, provides a fruitful environment for establishing a theory of tendering with an associated body of knowledge. A precursor activity is the design of a framework to study these interactions.

Keywords: Bayesian Belief Nets, bidding, contractor selection, decision analysis, ontology.

INTRODUCTION

The literature devoted to the general theory and practice of tendering appears to be sparse but those that have been published describe an unsatisfactory situation. (The term ‘tender’ and reciprocal term ‘bid’ will be used to reflect current usage.) Sturgess (2004) reflects on the absence of professional literature and asserts the practice of bidding to be akin to a black art. Published discussions from the construction industry are more numerous but similar sentiments of the “murky world of tendering” can be found (Hughes 2004). The acquisition of a product, services, a facility to be designed and built or a facility to deliver a service regime, require different forms of specification and evaluation. Wong et al. (1999) consider the shift from ‘lowest-price wins’ to ‘multi-criteria selection’ in an attempt to secure best value for money. As their focus is on the prospective construction of a facility, the multiple criteria generally reflect issues of Due Diligence (analysed with multivariate discriminant analysis). In contrast, a contract for the delivery of ongoing services (from a building) is very different to a building construction contract (Hughes 2002). The selection of the contractor moves beyond issues of Due Diligence to encompass the performance of multiple service level criteria. As these criteria are contractual terms, the methods of assessment must be robust. However, where such multiple criteria decision analysis is required to assess performance options, Bana e Costa et al. (2002) offer detailed criticism of the administrative use of published criteria weights and the resulting absence of mathematical meaning in “additive” multi criteria models. Runeson and Skitmore (1999) conclude that tendering theory falls uneasily between game theory, decision theory and auction theory and as such, is not based on a sound theoretical framework. The consequence of these findings may explain the paucity of activity to expand the theory of tendering. It is therefore valuable to describe the

1 a.white@adfa.edu.au

existing components of tendering theory and practice, applications which lack mathematical efficacy and potential methodologies which could provide a new direction for clients requiring an ongoing service delivery from their facilities. The development of a comprehensive theory may be difficult. Multiple criteria decision aids and associated methods are indicative of an art rather than a science (Keeney 1992, Hammond et al. 1999 and Checkland 1981). Rothkopf (1994) offers suggestions for further developments to bring bidding theory closer to the decision making process. With this background, an analytical framework will be introduced to aid discussion.

**DECISION MAKING AIDS – THEORY AND PRACTICE**

A global chorus emphasises that multi criteria decision aids support the decision making process and are not tools for conclusive selection (Hammond et al. 1999). For public sector policy makers, this singular issue needs to be addressed in current policies and procedures guiding the tender analysis process.

Belton and Stewart (2002) provide a thorough review of current (math.) techniques and practices. An equally valuable document is published by the UK government (DTLR 2001). While DLTR explores a wide range of techniques, Klein and Chapman (1996) provide insights into some of their underlying methodologies and procedural mathematics which if not understood (by policy makers) can lead to naïve and incorrect applications of the techniques.

Tendering theory employs various definitions of criteria across the procurement phases (compliance and conformance, differentiation, and risk assessment). Consequently these phases require different derivations of criteria weight. Public policy makers need to clearly identify these process phases and the different regimes of criteria and their relative weights in the tender documentation.

The following extract typifies public sector procurement guidance:

> Evaluation of suppliers should be based on the relative importance of each criterion. There are a variety of methodologies that can be used to achieve this. The most commonly used is to weight criteria according to their relative significance using a rating scale. (DFAT 2005).

Missing is the distinction between ordinal weights, where only the rank is considered (the weights suggest that one criterion is more important than another), and cardinal weights, where the value of each weight against others represents its relative importance for the decision maker. Belton and Stewart (2002) extend the definition of “ordinal” to “categorical” such as a semantic scale (very important, moderately important, etc.) The DFAT (2005) extract makes deliberate use of the term ‘relative’ and it is open to conjecture as to whether government decision makers are sensitive to the distinction between ordinal and cardinal weights and the potential to mix these weights in a multiple criteria decision making model.

The most intuitive method for combining a bidder’s performance at each criterion is to multiply the (normalised value) performance score by the respective criteria weight (Weighted Sum method). The weights must represent proportional preferences so that a criterion with a weight of 8 is twice as important as a criterion with a weight of 4. The resulting value function indicates the goodness of the alternatives. A higher rank implies a higher value alternative. Two mathematical processes are in operation. The first considers the aggregation of the judgements with regard to each criterion and
each alternative and the second considers the aggregation rules for ranking the alternatives. The question needs to be asked as to how much of this information is shared with potential bidders. As an example, AusAID (2004) advises:

*Preferred tenderers are selected on the basis of value for money. In the majority of instances, value for money is assessed based on a total score that combines a technical score (against weighted selection criteria) and a price score using a like-for-like price assessment of the financial proposals.*

Bana e Costa (2002) clearly states the practical implications when referring to the obligation to advise potential bidders of the evaluation criteria and their relative weights or at least the order of “importance” of the criteria prior to the submission of their bids. Firstly, the value of the weights should not change after they have been published (a potential legal issue). Secondly, how can the “weights” be elicited without reference to the actual criteria performances offered by the alternatives? That is, if weights have not been determined by reference to the total variation of submitted performances on all criteria, then the “weighted sum” is “theoretically incorrect and has no mathematical meaning in the framework of the additive model”. Klein and Chapman (1996) provide the mathematical reasoning. Bana e Costa (2002) continues with the observation that these popular additive methods using direct weighting methods (intuitive “importance”) are theoretically incorrect. Indeed, this brings into focus the distinction between criteria weights used to *filter* alternatives for conformance and compliance, and criteria weights used to *distinguish* alternatives passing the compliance and conformance filter.

There is no argument with criteria weights being used to advise bidders of the relative “importance” of criteria and the use of those published (and unalterable) weights to filter submissions for conformance and compliance. However, associating criteria weights with a specification is not a statement of “importance” but rather, is a statement of *tolerance*. That is, the specification reflects a singular design regime and the criteria weights (numeric scale or semantic scale of: mandatory, desirable, etc.) are statements of acceptable design tolerances.

An issue for public sector policy makers occurs precisely at the point where criteria weights are multiplied by a performance value or utility. The multiplication process is the turning point which brings the analysis into the realm of Multi-Attribute Value/Utility Function Theory (DLTR 2001). The Simple Additive Weighting (where performance scores are multiplied by criteria weights) can be expressed as:

\[ V(a_i) = \sum_{j=1}^{J} (w_j v_j(a_i)) \]

where:

- \( w_j \) is the weight of criterion \( j \)
- \( v_j(a_i) \) is the value of alternative \( a_i \) in criterion \( j \)

However, the efficacy of this procedure is maintained only if:

1. The criteria are preferentially *independent* (For example: Vehicle Mass and Fuel Efficiency are *dependent* criteria.),
2. The decision maker has an exact understanding about the utility of the performance scores of the alternatives and the weights of the different criteria (problematic if this is done prior to the assessment of the alternatives), and
3. The weights:
   a. are scaling constants (actual measured performances are converted through a Value Function to a scale of [0,1]. The weights change the relative value of all the [0,1] criteria value scales).
   b. have been derived from the actual performance ranges of the alternatives (Bana e Costa 2002 and Klein et al. 1986).
   c. are derived from a process of trade-offs which acknowledges complete compensation. (For example, how much am I prepared to trade-off the performance envelope of an aircraft with the electronic surveillance measures? More importantly though, is the notion of complete as opposed to partial compensation.)

The legal consequences need to be considered if public officials are unable to demonstrate that these principles have been maintained.

As the criteria transition into the next evaluation phase, their relative weights are likely to change. Consider a situation in which the specification associated the criterion Fuel Efficiency with an “importance” rating of Mandatory. That is, the fuel consumption had to be very close to 7 l/100km. The Mandatory requirement implies minimal acceptable performance tolerance. Once tenders had been opened, it was clear that all bidders were offering vehicles with a fuel efficiency of 7 l/100km. From then on, Fuel Efficiency has no value in differentiating the vehicles and consequently has a relative criterion weight of zero during the evaluation phase. For public sector policy makers, there is a need to remove the requirement for vertical assessment of bids (each bid is evaluated in isolation from all other bids) and introduce controlled horizontal evaluations in order to compute the relative criteria weights based on submitted performance data. Contrast this procedure with some prevailing (mandatory) government requirements:

In evaluating submissions the submission evaluation committee needs to ensure it rates the submissions against the evaluation criteria rather than other submissions (DFAT 2005).

“Rank reversal” is a weakness of the Weighted Sum method (Gelderman and Rentz 2000). This can be illustrated when one tenderer performs extremely well on low weight criteria and another performs poorly on high weight criteria (see Table 1).

| Table 1: Example of Rank Reversal – SAW. (Weight Range: 0-1, Score range: 0-100) |
|-----------------|-----------------|-----------------|-------------|
|                | Criterion 1     | Criterion 2     | Aggregate   |
|                | Weight = 0.7    | Weight = 0.3    | Score       |
| Bid            | Score          | Score          | Score x Weight =       |
|                | x Weight       | x Weight       | = Score x Weight       |
| Case 1         |                |                |                |
| A              | Score = 60     | Score = 60     | 60 ✓          |
|                | Score x Weight = 42 | Score x Weight = 18 |       |
| B              | Score = 48     | Score = 70     | 54.6          |
|                | Score x Weight = 33.6 | Score x Weight = 21 |       |
| Case 2         |                |                |                |
| A              | Score = 10     | Score = 60     | 25            |
|                | Score x Weight = 7 | Score x Weight = 18 |       |
| B              | Score = 8      | Score = 70     | 26.6 ✓        |
|                | Score x Weight = 5.6 | Score x Weight = 21 |       |

The Weighted Sum (SAW) method is just one model and Belton and Stewart (2002: 9) argue that the meaning of “weight” is model-dependent. However, information about the decision model is rarely published with the RFT. The discussion can be
expanded by summarising two schools of though in the field of Multi Criteria Decision Aids (MCDA). The “Anglo-Saxon / Teutonic” school (also referred to as the “American” school) is based on utility functions which are weighted and aggregated in order to solve the multi criteria problem. The Simple Additive Weighting method fits into this school of thought. The “Latin / Gallic” school (also referred to as the “French” or “European” school – Outranking techniques) analyses the underlying preferences on a pair-wise (alternatives) basis. Outranking assumes that preferences are not constant in time, are not unambiguous, and are not independent of the process of analysis (Brans et al. 1986).

The issue of “price” as a criterion in a multiple criteria decision aid requires consideration. If price has a high relative weight then there is little value in designing a decision aid model. Selection is based on the lowest price. In the context of benefit/cost analysis, the price is the cost, and the service delivering criteria are the benefits to be derived by the purchaser. Price is a good benchmark for undertaking trade-off studies for the determination of relative criteria weights. However, once used for this purpose, price should be removed from the decision aid model. Bids can then be assessed on total service delivery performance with respect to price. The Australian Department of Defence maintains this separation, in a practice known as the “two envelope” procedure. The price is submitted in a separate envelope and is not considered until the technical evaluation is completed. This means that only indicative cost based criteria can be used as benchmarks for trade-off studies to elicit criteria weights.

**JUDICIAL REASONING**

Periodically, judicial findings have provided elements of certainty from which practices have changed. In Hughes Aircraft (1997), the bidders were not informed of the changes in the weightings of the criteria. However, of relevance for future policy makers was the finding that the process leading to the award of the contract was governed by a preliminary “process contract” (see also Ron Engineering (1981)), the principal terms of which were contained in the RFT. An additional finding concerned an implied term (as a matter of law) that a public body must act fairly with a tenderer in the performance of a process contract. In this context, the potential misapplication of multiple criteria decision aids is emerging as an issue in disputes arising from the “process contract” and referred to adjudication or similar forum (author’s personal experience). Consequently, agencies are required to ensure that the tender process, particularly tender evaluations are consistent with the tender documents provided to tenderers, and that the basis of the evaluation is transparent on the face of the tender document. Unfortunately, the issues of mathematical efficacy raised by Bana e Costa (2002) have not been addressed. There may be an evolving realisation that multiple criteria decision aid mechanisms employing a performance or preference score multiplied by a (pre-evaluation) published criterion weight may have no place in public tendering. Anecdotally, some Australian government agencies have removed criteria weights from comparative tender evaluation.

**AUCTIONS, GAMES AND BAYESIAN BELIEF NETS**

Raiffa et al. (2002) place Game theory and Decision Analysis in context, where:

a) *Decision Analysis* is a prescriptive approach of how an analytically inclined individual should/could make wise decisions.
b) **Behavioural Decision Theory** is a descriptive approach on the psychology of how certain individuals do make decisions.

c) **Game Theory** is a normative approach of how groups of ultra-smart individuals should make separate, interactive decisions.

d) **Negotiation Analysis** is an integrated approach on how groups of reasonably bright individuals should and could make joint, collaborative decisions.

Dick (2005)\(^2\) firmly places tendering into Game theory.

“Because these competitions are governed by formal and informal rules and goals, they can be modelled as a game.”

The client sets the rules and bidders interact with their own strategies within these rules. The client seeks to provide detailed information and concurrently, the bidders seek to provide the best value solution. Such symbiotic relationships with variable outcomes fit uncomfortably with traditional multiple criteria decision aids. The resulting interactions produce at least six gaming environments. 1) Strategic Auction (lowest price) – where there is a congruence of understanding between the client and the bidders. 2) Beauty Contest (non-price focus) – where there is the same congruence of understanding and similar solutions are on offer. 3) Business Model Competition – where bidders are offering different solutions to a knowledgeable client. 4) Winner’s Curse – the winner always loses money because each bid is a different guess about the client’s needs and the winner’s guess is also wrong. 5) Arranged Marriage – where the client does not know what it wants but recognises the problem and 6) Blind Date – where the client does not recognise that bidders cannot focus on the ill-defined requirements.

Montibeller et al (2001) discuss a formalisation of the breadth of decision processes later described by Raiffa et al. (2002). Of interest is the inclusion of Bayesian Belief Nets (BBNs) which recognises that the composition of a tender and its evaluation by the client is focussed on a future and relatively unknown relationship. BBNs model problems that involve uncertainty and inference (Fenton 2001). In addition to statements about the probabilities of events, the bidder or the purchaser knows some evidence (observed) and needs to infer the probabilities of other events that have not as yet been observed. This is the essence if high risk tendering.

**A CONCEPTUAL FRAMEWORK FOR MODELLING THE TENDERING DOMAIN**

A framework is proposed to study the synthesis of the many bodies of knowledge contributing to public tender theory and to analyse the processes (see Figure 1).

Figure 1: A framework for the study of the public sector tendering domain

Part 1 distinguishes competitions searching for a capability based on detailed requirements and competitions searching for a solution (involving the definition of requirements). Part 2* was defined by Dick (forthcoming) and describes the evolution of specific Games within the tendering environment. Part 3 traces each Game through its evaluation phases. Phase 1 filters bids for conformance and compliance. Phase 2 seeks to differentiate the filtered bids. Criteria weights are derived from the actual performance ranges of the filtered bids (Bana e Costa 2002 and Klein 1986). Phase 3 is an interlocutory phase which seeks to understand the future environment. Risk is a relative measure. In the context of bid evaluation, it can be expressed as the cost of translating the performance of each short-listed supplier from their offered performance level up to the theoretically perfect performance level as described in the specification. That is, closure of the “performance gap”. This is a theoretical exercise modelled in a BBN environment as potentially, none of the tenderers may achieve the specified performance zenith. This interlocutory process has each tenderer defining the BBN and exposing the performance gap risks. Not all games will have all three evaluation phases. Much of the current government guidelines are restricted to Phase 1 of a Strategic Auction (DFAT 2005). Phase 2 cannot occur if the criteria weights are those published with the RFT which (by regulation) cannot be later changed, even though the criteria weights are serving a different purpose (Hughes Aircraft 1997). Phase 3 offers a formalisation (using BBNs) of the conventional negotiations phase with short-listed tenderers. However, it is a formalisation about the future and its risks. Further research may indicate that Phases 1 and 3 alone can provide a viable evaluation framework.

TENDERING DOMAIN ONTOLOGY AND ANALYSIS

While the proposed framework is merely a convenience for studying tendering theory and practice, the lack of a tendering ontology compounds the problems of quality
management. The process of (construction) contracting has long been recognised as an environment based on interdependence and uncertainty (Tavistock 1996). Access to timely information has been addressed with the availability of project extranets. Another key (remaining) issue is the development of a tendering domain specific ontology.

At its simplest, ontology is a classification system. More specifically, ontologies are agreements about shared conceptualisations. These include conceptual frameworks for modelling domain knowledge, content-specific protocols for communication among inter-operating agents, and agreements about the representation of particular domain theories (Gruber 1993).

The building construction industry (design) domain exists in the form of the Industry Foundation Classes (IFC) (AIA 2001). Kayed and Colomb (2002) apply ontology to the conceptual structures of the tendering domain. Their application uses a Conceptual Graph (CG) for knowledge representation. The effectiveness (congruence with the game) of a given set of contract documents (specification, agreement, etc.) within the tendering domain needs consideration. Governatori (2005) uses Deontic and Defeasible Logic to transform the logical form of these documents into machine readable rule notation using an XML based rule representation language RuleML. Grosof (2004) uses RuleML with ontologies to produce SweetDeal, an approach to aid automated creation, evaluation, negotiation and execution of contracts. Daskalopulu (1999) discusses the use of Petri-nets for assessing contract document performance. Petri-nets are tools for modelling dynamic systems (Peterson 1981 and Purvis 1998). Workflow management systems such as SAP R/3 and BaanERP base their modelling language on Petri-nets. Farrell et al (2004) have extended the technology of contract monitoring with the use of Event Calculus. They defined an ontology to capture aspects of Service Level Agreements for the purpose of state tracking and performance monitoring. Their ontology is formalised as an XML-based language called CTXML (contract tracking XML).

**CONCLUSION**

Several existing theories and bodies of knowledge have the potential to meld into a singular theory of tendering and to generate a tendering specific body of knowledge. A framework for studying the tendering environment will assist the definition of a tendering ontology. This is vital as any decision-making activity is enjoining technical, social and cultural influences. The use of multiple criteria decision aids may be restricted until they can be robustly enjoined with prospective risk evaluation procedures such as BBNs. However, the use of criteria weights beyond a preliminary filtering phase may be problematic. Neither the client nor the tenderer can perfectly predict the future but they do share a common voyage (alliance). A significant outcome of a tendering body of knowledge will be the empowerment of the legal profession to guide public sector policy makers from a defensive posture to one which is sympathetic to this voyage, the decision making phases and the game at hand.

**REFERENCES**


Hughes Aircraft Systems International v. Airservices Australia (1997) 146 ALR 1


**TERMINOLOGY**

MCDA Multiple Criteria Decision Aid

RFP Request for Proposal

RFT Request for Tender

SAW Simple Additive Weighting