Munich Social Science Review, New Series, vol. 8(2024)

How to Define Cherries to Pick in Industrial Procurement[1]

Gregor Berz
Institut für Angewandtes Mechanismus Design (IFAMD),
Kutaisi International University (KIU) and Center of Conflict Resolution (CCR)
gregor.berz@ifamd.de

Evguenia Shprits
Institut für Angewandtes Mechanismus Design (IFAMD)
evguenia.shprits@ifamd.de

Thorsten Velten
Deutsche Bahn AG
thorsten.velten@deutschebahn.com

Abstract: We introduce a new approach to the bundling of lots in procurement auctions and other business tendering processes. Based on a Competition Matrix filled with provider historical prices, a division of the demand into partial bundles ("*Cherries*") is determined, which aligns with the providers' specializations according to their price patterns. In a world with perfect information, the IFAMD Cherry Picker would be evaluated on cost information from the providers and not based on prices. As a result, Cherry picking from the side of buyer is ensued: Cherries are those items which can be procured with especially lower prices, due to the possible specialisation of the providers. An assessment of whether this specialization is genuine and due to production costs and other factors influencing the price or motivated by tactical pricing due to implicit collusion is not considered at this stage. It is recommended to conduct this strategic examination before using

[1] Documentation report of the "Cherry picking" sub step within the "Game Theory with Big Data from ProLake" project of the Institute for Applied Mechanism Design (IFAMD) GmbH for Deutsche Bahn AG in December 2022.

ISSN 0170-2521
ISBN 978-3-88278-317-9
www.accedoverlag.de

the Cherries identified by the IFAMD Cherry Picker to define lots in a business tendering process.

1. Introduction

The bundling of lots is one of the procurer's most crucial and challenging decisions, which is widely used in practice and underrepresented in the theoretical literature (according to Schoenherr and Mabert 2011).

The number, the size and the shape of lots directly influences the number of bidders, the bidders' behavior and potential collusion in the long and in the short run (Grimm et al. 2006, Chen and Li 2018). Moreover, bundling can heighten bargaining power for the buyer (Schoenherr and Mabert 2006). In addition, bundling several items together can directly impact cost components, such as the purchase price and operations cost, generating economies of scales for the buyer as well as for supplier (Schoenherr and Mabert 2011). According to Schoenherr and Mabert (2011), the bundle building is associated with a goal to avoid Cherry picking from the side of suppliers: suppliers only submitting bids on the most attractive items and avoiding less desirable ones. Finally, the attractiveness of a lot is related to its reachable price in a tendering process and thus to the level of competition which is reduced, if bidders focus on different Cherries each. The measure of choice to avoid this bidders' strategic demand reduction is forced bundling as described in Berz (2015: Section 8.3).

Despite the wide usage in practice, existing economic analysis provide only limited guidance when it comes to deciding the number of lots in which a supply contract should be divided, and deciding about their sizes. According to Grimm et al. (2006), the basic rule "The number of lots should be smaller than the expected number of participants" emerged from the literature. Beyond this basic rule, the discussion so far has not clarified how to choose the right number and size of lots. In practice, typically items are bundled together if they rely on similar raw materials or similar production processes, or if they have similar applications or technical requirements (Schoenherr and Mabert 2006). Similarity between complicated project types is an important consideration in project bundling: Qiao, Fricker, and Labi (2019) present a methodology to quantify the similarity between different project types based on their constituent pay items. But in order to break off collusive behavior, only the synchronization of supplier decision is necessary for forced bundling as described in Berz (2015: Section 8.3).

We introduce a method for bundling based on the idea of *cherry picking* from the buyer's side: the items which can be procured with especially lower prices, due to heterogeneity in production costs and other factors influencing the price, by one specific supplier, would be bundled together in one *Cherry*. Because we assume that the production costs are private information, the bundling decision relays on the historical data of the price patterns from the previous demand. But before using this simplification it is essential to assume that collusive behavior on the bidders' side does not occur.

In the next section, section two, we introduce our model. Section three contains two examples. Section four concludes.

2. The Model

2.1. Price Matrix

First, for an upcoming demand we define *Condition Atoms*. These are the smallest units of segmentation within the relevant demand that are to be priced in the offers from providers or within contracts. The Condition Atoms make up the rows of the *Price Matrix*. From historical data pools, such as the DB ProLake, prices for these Condition Atoms from various providers are now to be identified. For each provider from which prices are found, a column is created in the Price Matrix. If these prices come from different data sources (e.g., predecessor business tenders of the same material or service), then consistency, i.e., the comparability of the prices, should be ensured. Prices from different time periods can potentially be adjusted to each other using a price adjustment logic. If multiple prices are available for the same *Condition Atom* for a provider, it is advisable to use the more recent price when in doubt.

Condition Atoms are generally structured based on several *Bundling Dimensions*, such as material numbers, regions, different demand carriers, or delivery points, etc. For the application of the algorithm described below, it is recommended to select one of these Bundling Dimensions and aggregate the conditions across all other Bundling Dimensions. This corresponds to the philosophy that "the game theorist does not bundle bottom-up but de-bundles top-down". If good Cherries are found in this way for more than one bundling dimension, they can be multiplied with each other afterward.

Without loss of generality, it is assumed that the rows of the price matrix represent *Condition Atoms*. If they are aggregations in the sense of the previous paragraph, then these aggregations will take on the role of *Condition Atoms*.

The *Price Matrix* is denoted as $\mathbf{P}_{m\times n}$, where m is the number of *Condition Atoms*, and n is the number of providers. The matrix element $p_{i,j}$ corresponds to the price found by bidder j on *Condition Atom* i. If no price was found for a provider j on *Condition Atom* i, then we set $p_{i,j} = 0$. Next we define the vector of the best prices for all *Condition Atoms* $i = 1, \ldots, m$:

$$p_i^{min} = \min\{p_{i,1}, \ldots, p_{i,n} | p_{i,j} \neq 0\} \qquad (1)$$

2.2. Specialist Matrix

Let $\varepsilon \geq 0$ be a *Specialist Tolerance*. In general, an epsilon vector $(\varepsilon_1, \ldots \varepsilon_m)$ is defined, depending on ε. As an example, this may be:

$$\varepsilon_i = \text{Covariance}\,(p_{i,j}, j = 1, \ldots, n) \cdot \varepsilon \qquad (2)$$

In the current paper we have always set $\varepsilon_i = \varepsilon$ for all $i = 1, \ldots, m$. In any case, the *Specialist Matrix* $\mathbf{S}_{m\times n}$ is defined as follows:

$$s_{i,j} = \begin{cases} 1 & if \quad 0 \neq p_{i,j} \leq p_i^{min} \cdot (1 + \varepsilon_i) \\ 0 & else \end{cases} \qquad (3)$$

Each entry $s_{i,j} = 1$ in the *Specialist Matrix* means that the price of provider j in the respective *Condition Atom* i deviates by less than $\varepsilon_i \cdot 100$ in % from the best price. Thus, the provider is considered a specialist in this *Condition Atom*. In the case of $\varepsilon = 0$, each *Condition Atom* has only one specialist, namely the provider with the best price. As ε increases, the *Specialist Matrix* fills up monotonically until, finally, with a sufficiently large ε, the *Specialist Matrix* converges and becomes the *Competition Matrix.*

2.3. Cur-or-Join Matrix

In this section, first we define the *Join Matrix* $\mathbf{N}_{m\times m}$ as follows: for each pair of *Condition Atoms* i and j, $n(i,j)$ is defined as the relative proportion

of providers that are either specialists in both *Condition Atoms* or not specialists in both *Condition Atoms*.

$$n_{i,j} = \frac{1}{n}\sum_{k=1,\dots,n} \begin{cases} 1 & if \quad s_{i,k} = s_{j,k} \\ 0 & else \end{cases} \qquad (4)$$

To come to the *Cut-or-Join Matrix*, we define a trimming parameter $\delta \in [0,1]$. Two *Condition Atoms* i and j are considered "to be joined" if $n(i,j)$ is greater than or equal to δ. Otherwise, the two *Condition Atoms* are considered "to be cut". The Cut-or-Join Matrix $\mathbf{C}_{m\times m}$ is defined as follows:

$$c_{i,j} = \begin{cases} n_{i,j} & if \quad n_{i,j} \geq \delta \\ 0 & else \end{cases} \qquad (5)$$

2.4. Cherries determination and assessment

The idea behind the Cut-or-Join Matrix is that the sought-after "*Cherries"* are the connected components of the adjacency graph of the Cut-or-Join Matrix. The adjacency graph of a square matrix is the graph that results when you interpret the rows or columns as nodes (which, in this context, are the Condition Atoms). Two nodes or Condition atoms, i and j, are connected by an edge if the matrix entry at position (i,j) is not equal to 0.

The determination of connected components in the adjacency graph of a matrix is a mathematical problem for which there are standard solution algorithms. For the programming language R, which we have been using in our discussions, such a solution algorithm can be found in the open-source library *igraph* under the function name *components*.

In any case, the connected components of an adjacency graph are unique, meaning that the result of identifying the Cherries does not depend on which solution algorithm you use to determine the connected components.

The number of identified Cherries (connected components) depends on ε and δ.

The trimming parameter δ ensures that when $\delta = 1$, each provider in all Cherries can either offer all Condition atoms as specialist or none Condition atom as specialist. However, it is suspected that the strict interpretation of $\delta = 1$ leads to the discovery of good Cherries only rarely. With a tolerance of $\delta < 1$, but still close to 1, Cherries can be found where individual

providers may offer them as specialists with partial coverage. To assess the usability of Cherries, we calculate the *Coverage Rate* of bidder j for Cherry R as the relative proportion of *Condition Atoms* in R that bidder j offers. More precisely, the *Coverage Rate* is considered to be weighted with the best prices p_i^{min}:

$$\text{Coverage Rate}_j^{R,\varepsilon} = \frac{\sum_{i \in R} p_i^{min} \cdot s_{i,j}}{\sum_{i \in R} p_i^{min}} \qquad (6)$$

It's important to note that the Coverage Rate also depends on ε since the Specialist Matrix with entries $s_{i,j}$ depends on ε. From a procurement law perspective, one would probably have to consider the full *Competition Matrix*, which occurs when ε is large enough ("$\varepsilon = \infty$"). However, from a strategic standpoint, especially when dealing with historical data, it's essential to question the tolerance level for including a bidder who is not a specialist in the definition of lots through Cherries.

In a business tendering process, comparing bids from bidders with different *Coverage Rates* can be achieved using price patterns that are predefined identically for all bidders (as for "gaps in the Competition Matrix"). The decision of, in a business tender, at what Coverage Rate < 1 a bidder is allowed to bid on and may win a lot (defined as a Cherry) should be made strategically based on the competition situation of the tender. In cases involving procurement law, the exclusion of bidders from lots for which they have a Coverage Rate > 0 but below a certain threshold needs to be legally clarified.

The Specialist Tolerance ε is not a parameter chosen by the user but is increased from a lower value in an iterative process until a stable distribution of Cherries is achieved. The range of the ε value interval in which a Cherry does not change is called its "*Cherry Stability"*. As an additional metric for evaluating the identified Cherries, we suggest calculating the average of all edge values, considering all missing edges within the Cherry. This can be referred to as the "*Cherry Quality"*:

$$\text{Cherry Quality}^{R} = \frac{\sum_{i,j \in R} c_{i,j}}{\text{Count}\,(i \in R)^2} \qquad (7)$$

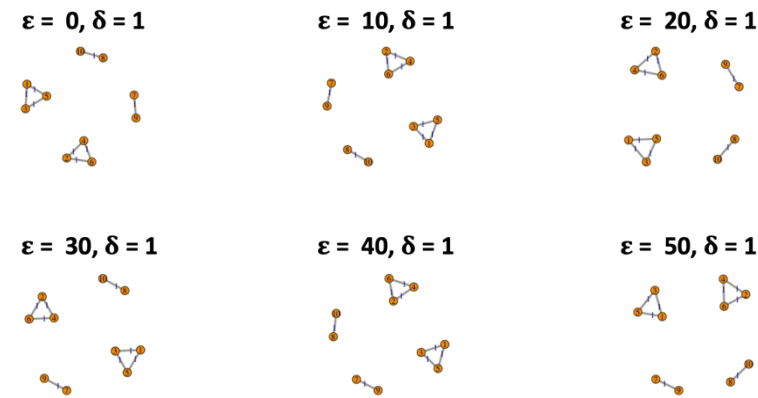
By setting $\delta = 1$, it's indeed a fact that the Cherry Quality can only be 1. However, when $\delta < 1$, it's important to consider Cherry Quality to avoid

diluting the strategic bundling with "soft" Cherries. This emphasizes the need to maintain a balance between the strictness of the condition and the quality of the Cherries in that case, ensuring that the chosen Cherries align with the strategic objectives of the bundling process.

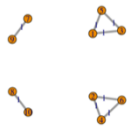
3. Examples

The first example is a synthetic *Price Matrix* that exhibits classic W-M-shaped price patterns among four providers based on ten conditioning factors. The second example consists of 100 randomly selected spare parts with prices from eight vendors, derived from real data of the DB.

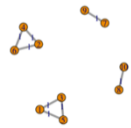
In the first example, already with $\varepsilon = 0$, the *IFAMD Cherry Picker* identifies four Cherries that remain stable until $\varepsilon = 80$. When $\delta = 1$, these Cherries break down into six at $\varepsilon = 90$ and into eight at $\varepsilon = 100$, only to merge back into one connected component at $\varepsilon = 130$:



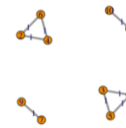
ε = 60, δ = 1



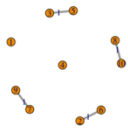
ε = 70, δ = 1



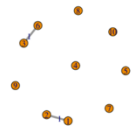
ε = 80, δ = 1



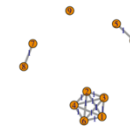
ε = 90, δ = 1



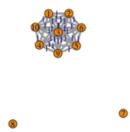
ε = 100, δ = 1



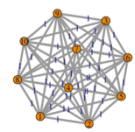
ε = 110, δ = 1



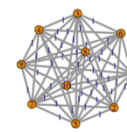
ε = 120, δ = 1



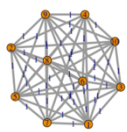
ε = 130, δ = 1



ε = 140, δ = 1

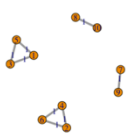


ε = 150, δ = 1

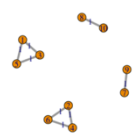


With δ = 0.7, a similar pattern emerges up to ε = 80, but at ε = 90, instead of breaking apart, they form two soft Cherries with Cherry Quality values of 0.53 and 0.88, and at ε = 100, they merge into one Cherry with a Cherry Quality of 0.35:

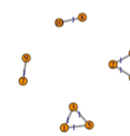
ε = 0, δ = 0.7

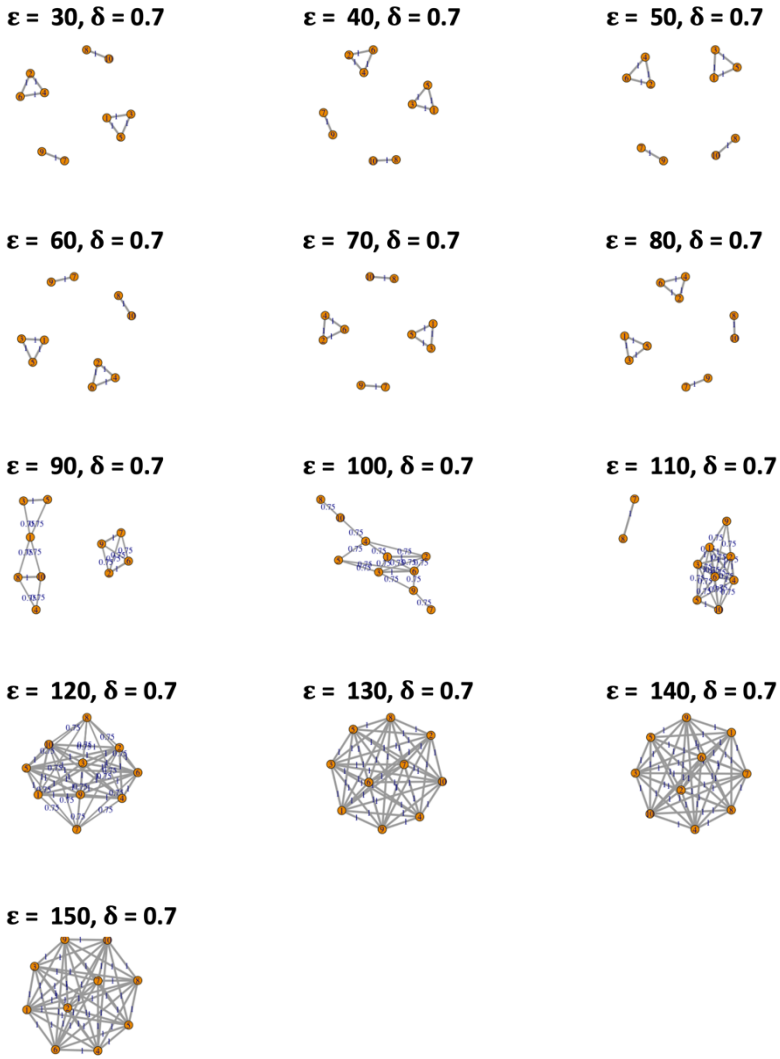


ε = 10, δ = 0.7



ε = 20, δ = 0.7

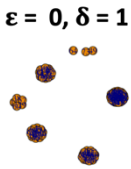


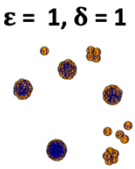


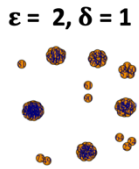
In the overall assessment, in the example 1, one would likely choose the four Cherries from the ε range between 0 and 80. However, in the real world, it's crucial to verify whether these Cherries indeed represent specialization or are the result of tactical pricing due to (implicit) collusive motivation.

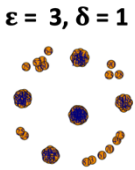
In example 2, with $\delta = 1$ and $\varepsilon = 0$, seven Cherries are identified because seven of the eight providers offer the best prices for some of the 100 spare

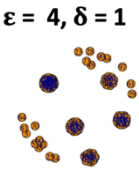
parts. However, as ε increases, these Cherries break down into smaller subsets, and none of them remain stable for an extended epsilon range.

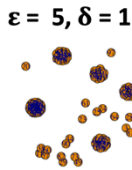


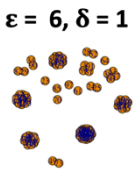


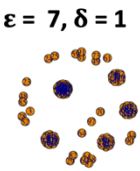


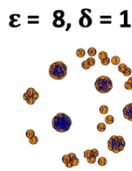


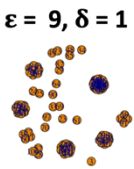


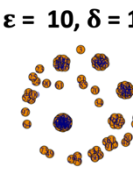




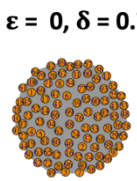


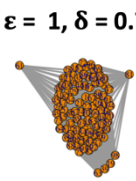


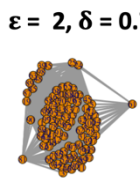


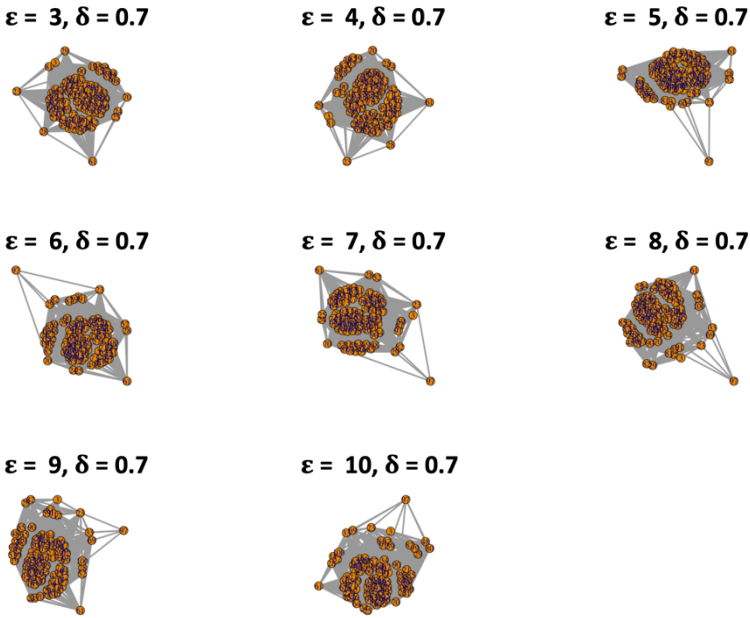


With δ = 0.7, all 100 spare parts are immediately grouped into a single Cherry, but its quality drops from 0.8158 at ε = 0 to 0.5477 at ε = 10:









In the overall assessment of example 2, no subset of the business forms a stable Cherry. The large number of separated subsets at δ=1 is not suitable for lot definition since there are many more subsets than eligible providers, potentially leading to implicit collusion among providers. In the real world, a decision was made for a larger lot where the approved providers had significantly different Coverage Rates, and the bid comparison was facilitated using price patterns predefined identically for all providers.

The plots were generated using the layout_with_fr function from the open-source *igraph* library in the R programming language, which applies the Fruchterman-Reingold layout algorithm.

These examples illustrate the complexity of the IFAMD Cherry Picker's behavior in different scenarios and the strategic considerations involved in choosing Cherries for lot definitions.

4. Conclusion

The IFAMD Cherry Picker is a method for dividing the demand to be tendered into bundled lots according to the specialization of the bidders. There are several empirical works which investigate how the existing

bundling practices influence the auction results. Estache and Iimi (2011), using evidence from water supply and sewage projects, show that there is a trade-off between competition in auctions and size of contracts and that it is important to account for the underlying cost structure for designing efficient auction mechanisms. In our model, we only use prices due to availability of data. The same model and calculations are feasible and free from the risk that prices could be influenced by implicit collusion or other bidding tactics if bidders' cost information is available.

Appendix (Glossary)

Adjacency Graph

Term from graph theory: For a square matrix, the adjacency graph is defined by the columns (= rows) of the matrix as nodes, and there is an edge between two of these nodes if and only if the corresponding matrix entry is not equal to zero. The matrix entry itself is considered as the edge weight. If the matrix is not symmetric, it is referred to as a directed graph. Here, we are exclusively considering adjacency graphs derived from symmetric Cut-or-Join Matrices.

Bundling Dimension

Conditions in a demand can be structured based on attributes, such as material numbers, regions, different demand carriers, delivery points, etc. All attributes belonging to a common category may serve as a Bundling Dimension.

Cherry

A subset of a business or upcoming demand in which the eligible providers can be clearly categorized into specialists and non-specialists.

Cherry Quality

The average value of all entries in the Cut-or-Join Matrix between the Condition Atoms of the Cherry. Cherry Quality indicates how compact or soft a Cherry is. With the trimming parameter $\delta = 1$, only a Cherry Quality of 1 can occur. For $\delta < 1$, Cherry Quality is of interest.

Cherry Stability

The range of the value interval for the Specialist Tolerance ε in which the Cherry does not change.

Condition Atom

The smallest units to be priced in offers from providers or in contracts for a given business or upcoming demand.

Coverage Rate

The Coverage rate of a provider for a Cherry is the proportion of the Cherry for which prices from the provider are available.

Cut-or-Join Matrix
A symmetric matrix of dimension ([condition atoms] × [condition atoms]), in which each nonzero entry indicates that the two respective condition atoms are generally intended to be connected within a Cherry.

Join Matrix
A symmetric matrix of dimension ([condition atoms] × [condition atoms]), in which each entry on a scale between 0 and 1 indicates the extent to which the two respective Condition Atoms are intended to be connected within a Cherry.

Predecessor tender
Historical tender of predecessor demand that is currently up for bidding.

Price Matrix
A Competition Matrix in which the entries are prices from the providers corresponding to the rows of the matrix.

Specialist
A provider whose price for a Condition Atom is closer to the best price than the Specialist Tolerance.

Specialist Matrix
A $0-1$ matrix of dimension ([condition atoms] × [providers]), where each entry of 1 indicates that the respective provider is considered a specialist for the corresponding Condition Atom.

Specialist Tolerance
Width of the tolerance range around the best price pattern, within which any provider whose price for a Condition Atom falls within this range is considered a specialist for that Condition Atom.

References

Berz, G. (2015), *Game Theory Bargaining and Auction Strategies*. London: Palgrave-Macmillan.

Chen, Y., and J. Li (2018), "Bundled procurement," *Journal of Public Economics* 159: 116–127.

Estache, A., and A. Iimi (2011), "(Un) Bundling Infrastructure Procurement Evidence from Water Supply and Sewage Projects," *Utilities Policy* 19: 104–114.

Grimm, V., et al. (2006), "Division in Lots and Competition in Procurement," *Handbook of Procurement,* Cambridge, UK: Cambridge University Press.

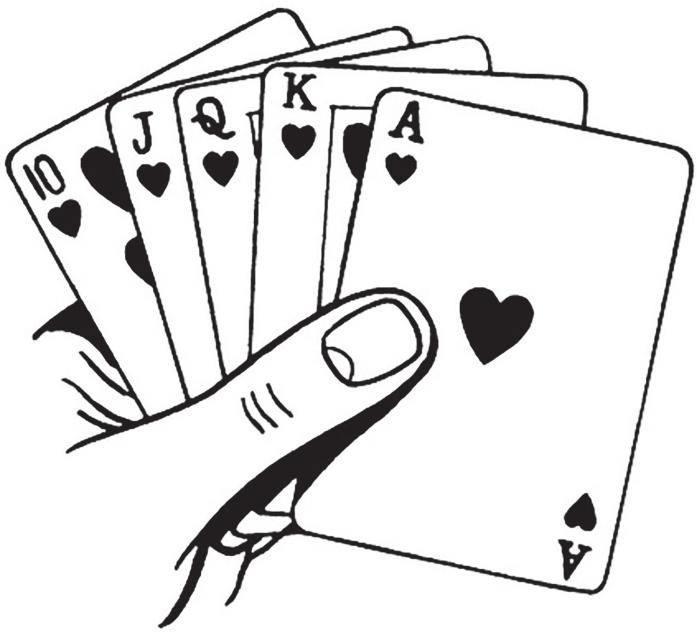
Qiao, Y., Fricker, J. D., and S. Labi (2019), "Quantifying the Similarity between Different Project Types Based on Their Pay Item Compositions: Application to Bundling," *Journal of Construction Engineering and Management* 145: 56–97.

Schoenherr, T., and V.A. Mabert (2011), "An Exploratory Study of Procurement Strategies for Multi-Item RFQs in B2B Markets: Antecedents and Impact on Performance," *Production and Operations Management* 20: 214–234.

Schoenherr, T., and V.A. Mabert (2006), "Bundling for B2B procurement auctions: Current state and best practices," *International Journal of Supply Management* 3: 189– 213.

Munich Social Science Review

New Series, Volume 8 / 2024



ACCEDO

Verlag Holler, München

Munich Social Science Review (MSSR)
New Series, Volume 8/2024

Applying Game Theory by Those Who Manage

Edited by
Gregor Berz & Florian Rupp

ACCEDO
Verlag Holler, München

IFAMD
game theory insights

Munich Social Science Review (MSSR), *New Series* Volume 8/2024

Cover: Katharina Kohl
Cover image: CSA Images via Getty Images

Gnesener Str. 1, D-81929 München, Germany
www.ccr-munich.de, manfred.holler@accedoverlag.de

Print: Libri Plureos GmbH, Friedensallee 273, D-22763 Hamburg

Book orders can be placed with your local bookstore or with Amazon, etc.

ISSN 0170-2521
ISBN 978-3-88278-317-9

Munich Social Science Review (MSSR)
New Series, Volume 8/2024

Applying Game Theory by Those Who Manage

ISBN 978-3-88278-317-9
ISSN 0170-2521