EURO Journal on Decision Processes 14 (2026) 100067

Contents lists available at ScienceDirect

EURO Journal on Decision Processes

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

Preferential knowledge for multi-criteria decision making: Stability and consistency of decision rules and weights

Marko Bohanec [a,*], Nikola Kadoić [b], Nina Begičević Ređep [b]

[a] Jožef Stefan Institute, Department of Knowledge Technologies, Jamova cesta 39, SI-1000 Ljubljana, Slovenia
[b] University of Zagreb, Faculty of Organization and Informatics, Pavlinska 2, HR-42000 Varaždin, Croatia

ARTICLE INFO

Keywords:
Multi-criteria decision modelling
Preference knowledge
Method DEX
Decision rules
Weights
Stability
Consistency

ABSTRACT

This research investigates the variability of decision-making preferences, represented in terms of decision rules and criteria weights, in the context of the qualitative multi-criteria method DEX (Decision EXpert). We study the differences between decision rules acquired from different subjects (inter-personal differences) and from the same subjects at different times (intra-personal differences). We also assess the consistency of so-acquired rules and the ability of subjects to estimate the importance (weights) of criteria. The methodological approach consisted of two surveys among students, carried out about one and a half month apart. Four thematic areas were addressed in the questionnaires: selection of study programs, student success, car purchase decisions, and choices regarding everyday shopping venues. In both survey periods, participants were required to assess the importance of these criteria and to define decision rules according to the DEX method. The findings provide insights into the stability of decision-making processes among participants and in time. The results indicate a high variability of decision rules, both inter- and intra-personal. Intra-personal drift is lower than inter-personal differences, but not by much (three-quarters of the latter). The consistency of rules varied between small decision tables with clearly ordered criteria, where it was almost perfect, and large decision tables with less apparent preferential relations. Defining fully consistent decision tables turned out to be hard, indicating the need for automated consistency-checking tools. Criteria weights also drifted in time at the rate about 9% (user-provided weights) and 10–27% (weights assessed algorithmically from decision rules). The main contributions of this study are identified and quantified magnitudes of decision rules variability and consistency.

1. Introduction

Multi-criteria decision modelling (MCDM) is a decision-making technique that involves the use of models to evaluate a set of decision alternatives based on multiple criteria or objectives (Greco et al., 2016; Thakkar, 2021). MCDM is used in situations where the decision maker needs to balance the trade-offs between multiple, possibly conflicting criteria. MCDM typically involves three steps: (1) defining the decision problem and criteria, (2) identifying and evaluating the alternatives, and (3) synthesizing the results to decide. There are many MCDM methods that differ in how they represent criteria, evaluation/aggregation rules and alternatives, and how they acquire this information, which is often subjective, from decision makers. MCDM methods are typically named using acronyms, such as AHP, ANP, TOPSIS, MACBETH, PROMETHEE, ELECTRE, DRSA, DEX; see Greco et al., (2016), Thakkar (2021) and Kulkarni (2022) for overviews and more information.

In this study, we are particularly interested in aggregation/evaluation aspects of multi-criteria models. In order to evaluate alternatives, a vast majority of MCDM methods employ the weighted sum:

$$f(x_1, x_2, \ldots, x_n) = w_1 x_1 + \ldots + x_n x_n$$

Here, x_i and w_i denote numerical criteria and their weights, respectively. The larger the weight, the more influential the criterion. Weights $w_1, w_2, \ldots, w_n$ are often normalized so that their sum or maximum equals to some predefined number, typically 1 or 100. Generally, weights are subjective and need to be acquired from individual decision makers (Rezaei et al., 2021; Silva et al., 2021).

On the one hand, MCDM methods strive to obtain weights that represent decision maker's preferences as accurately as possible. A good example is the method AHP (Analytic Hierarchy Process) (Saaty and Vargas, 2012), which proceeds by asking the user to assess relative importance of pairs of criteria, using the scale from 1 (equal importance)

* Corresponding author.
E-mail address: marko.bohanec@ijs.si (M. Bohanec).

https://doi.org/10.1016/j.ejdp.2025.100067
Received 9 April 2025; Received in revised form 3 December 2025; Accepted 17 December 2025
Available online 18 December 2025
2193-9438/

to 9 (extreme importance of one criterion over other). On this basis, AHP calculates criteria weights and assesses the consistency of user's information.

On the other hand, weights are subjective. Not only that they differ between different decision makers, they can change ("drift") also with the same person due to changes in the decision context, changes of their preferences or just their inability to express their preferences accurately enough. Thus, the question is how well can we assess criteria weights and what inter- and intra-personal differences should we expect.

This study is aimed at answering these questions in general and in relation with the decision modelling method DEX (Decision EXpert) (Bohanec, 2022). DEX is a qualitative MCDM method. It is somewhat specific in that it uses qualitative criteria and decision rules. Variables that represent criteria in DEX models are not numeric, but discrete and symbolic, using words as their values instead of numbers. For example, the criterion *Price* can be assessed using three categories "high", "medium", "low", and *Technical characteristics* of some system can be assessed as "bad", "acceptable", "good", or "excellent". Consequently, in order to evaluate decision alternatives, DEX does not employ the weighted sum, but *decision rules* that take the general form:

$$\text{if } x_1 = v_1 \text{ and } x_2 = v_2 \text{ and } \cdots \text{ and } x_n = v_n \text{ then } f(x_1, x_2, \cdots, x_n) = v_y$$

Here, x_i are qualitative criteria and v_i are some categories taken from the corresponding value scales. Similarly, as with weights, decision rules are acquired from the decision maker and conveniently represented in terms of *decision tables* (see example in Table 1).

In this way, DEX incorporates (Bohanec, 2022) several concepts that are typical of expert systems (Jackson, 1998; Leondes, 2002): using qualitative (symbolic) variables, representing decision knowledge in terms of "if–then" rules, and emphasizing the transparency of decision models. Also, DEX can operate in the case of missing and uncertain input information, and emphasizes the explanation of results; while these aspects are typical for expert systems, they were not considered in this study.

In this study, we addressed the following research questions:

A. *Inter-personal differences*:
   A1: How much do decision tables acquired from different subjects differ?
   A2: How much do quantitative criteria weights acquired from different subjects differ?
B. *Intra-personal differences*:
   B1: How much do decision tables acquired from the same subject at different times differ?
   B2: How much do quantitative criteria weights acquired from the same subject at different times differ?
C. *Consistency*:
   C1: Are decision rules, formulated by subjects, preferentially consistent and to which extent?
   C2: How do consistencies of decision rules from the same subject at different times differ?
D. *Weight assessment*: How much do quantitative weights, formulated directly by subjects, differ from quantified weights inferred indirectly from decision rules defined by the same subjects?

This study builds on and substantially extends a previous preliminary study (Bohanec, 2023), which addressed a subset of the above research questions, investigated two thematic areas (car purchase and shopping) instead of four and involved a relatively small number of participants (34), who were questioned at different times and occasions – this was insufficient for a reliable and conclusive scientific analysis. The present study was conceived to extend the research questions, increase the number of participants and carry out a well-controlled experimental setup with precise time differences between the trials. The participants from the preliminary study were not involved in the current study, nor

Table 1
A table for acquiring decision rules for the evaluation of study programs.

	Match with interests	Employability	Complexity	Study program			
				unacc	accept	good	***excel***
1	***low***	***low***	**high**				
2	***low***	***low***	medium				
3	***low***	***low***	***low***				
4	***low***	medium	**high**				
5	***low***	medium	medium				
6	***low***	medium	***low***				
7	***low***	**high**	**high**				
8	***low***	**high**	medium				
9	***low***	**high**	***low***				
10	medium	***low***	**high**				
11	medium	***low***	medium				
12	medium	***low***	***low***				
13	medium	medium	**high**				
14	medium	medium	medium				
15	medium	medium	***low***				
16	medium	**high**	**high**				
17	medium	**high**	medium				
18	medium	**high**	***low***				
19	**high**	***low***	**high**				
20	**high**	***low***	medium				
21	**high**	***low***	***low***				
22	**high**	medium	**high**				
23	**high**	medium	medium				
24	**high**	medium	***low***				
25	**high**	**high**	**high**				
26	**high**	**high**	medium				
27	**high**	**high**	***low***				

was their data considered in the analysis.

2. Related work

In this section, we review research relevant to how people formulate and express preferences, how stable those preferences are, and how this relates to multi-criteria decision models such as DEX.

2.1. Human judgment and decision making

Understanding human problem-solving strategies and information processing is essential for modelling human decision making and simulating human judgments. The science of judgment and decision making is typically described through three interconnected theoretical perspectives: normative theories, which identify optimal decisions; descriptive theories, which examine actual behavior; and prescriptive theories, which aim to help people make better choices (Kahneman, 2003). These perspectives correspond to three major research streams: analyzing the decisions people make, describing their natural responses, and designing interventions to improve those decisions (Fischhoff and Kadvany, 2011).

Fischhoff and Broomell (2020) distinguished three essential elements of decision making: judgment, preference, and choice. Judgments involve predicting outcomes of decisions, preferences involve evaluating the importance of those outcomes, and choices integrate these components. There are two major criteria for evaluating judgments, accuracy and consistency, which form a central analytical basis for understanding human decision processes.

Accuracy and consistency do not necessarily coincide. Individuals may be accurate on some topics yet inconsistent on closely related ones, or they may demonstrate internal consistency while holding limited knowledge (Fischhoff and Broomell, 2020). This asymmetry highlights how sensitive human judgments are to contextual and task-related factors.

These foundational insights reveal why eliciting stable and internally consistent preferences from individuals is difficult, motivating empirical investigations into how such judgments manifest in structured MCDM environments.

2.2. Stability, drift, and contextual factors in human decision making

Sensitivity to task features can substantially influence judgments and their stability (Fischhoff and Kadvany, 2011). These features include individual differences, life-span changes in decision competence, the distribution of outcomes, the amount of reflection time available, and the level of task comprehension. Environmental and contextual factors play a similar role. For example, in the auditing domain, Santos and Cunha (2021) demonstrated that trust, time pressure, and task complexity influence both the effort applied and the resulting judgments.

Further evidence of contextual sensitivity comes from studies of question-order effects. Novella and Ramirez (2024) found that expectation and risk-aversion measures vary depending on the order of survey modules, indicating that subtle priming can shift judgment under uncertainty. Smoliński and Brycz (2024) studied the accuracy of economic judgement and found substantial individual differences in cognitive biases, challenging the notion of a universal, context-independent predictor of bias. Time pressure produces yet another type of instability. Edland and Svenson (1993) showed that under high time pressure, people shift decision strategies and overweight negative attributes, altering both the structure and the content of their judgments.

The dynamic nature of preferences is additionally reflected in how individuals interpret decision weights. Vohs and Luce (2010) demonstrated that small situational changes can alter how people value outcomes, supporting the notion of constructed rather than stable preferences.

Taken together, these findings consistently demonstrate that human judgment is dynamic and context-dependent, raising important questions about the stability of preferences when represented through explicit decision rules and weights.

2.3. Ability to report attribute weights and metacognitive limits

Another line of research examines how well individuals can introspectively report the factors influencing their decisions. Nisbett et al. (1977) argued that people have limited direct access to their cognitive processes; instead, they construct post-hoc explanations based on implicit causal theories.

This issue becomes critical in multi-attribute decisions requiring explicit weighting of criteria. Cash and Oppenheimer (2025) introduced the Knowledge of Weights (KoW) paradigm, a method for assessing metacognitive insight into attribute weights without requiring comparison to a correct answer. Their findings show substantial variability and miscalibration in people's beliefs about their own weighting processes.

Complementary evidence comes from Morris et al. (2025), who used computational modelling to compare participants' self-reported decision strategies with their actual choice processes. While some participants demonstrated high introspective accuracy, many did not, and substantial individual variability was observed across five studies. These findings challenge the notion that people are strangers to themselves, suggesting instead that individuals often know how they made their value-based choices.

Together, these studies suggest that although individuals often feel confident about their preferences, their ability to accurately articulate attribute weights is limited and unstable. This highlights the need for systematic quantification of inter-individual differences, controlled measurement of intra-individual instability, and empirical validation of weight-assessment methods. To address these empirical gaps, our study investigates differences across subjects and across time in decision tables and criteria weights, providing essential metrics on subjectivity, instability, and the reliability of explicit decision rules.

2.4. Dynamic MCDM and the position of DEX

Traditional expected utility theory assumes stable preferences and probability-based evaluations (Moscati, 2023), whereas prospect theory conceptualizes decisions as driven by subjective decision weights and constructed preferences (Tversky and Kahneman, 1992).

Research in multi-criteria decision making provides convincing insights into the dynamic and evolving nature of human preferences. For example, Ariely and Loewenstein (2000) showed that people adjust evaluations when the duration of an experience changes, indicating flexible weighting of attributes. Decision stressors, such as complexity, information overload, time pressure, and uncertainty, can degrade judgment quality (Phillips-Wren and Adya, 2020). Emotional and risk-attitude factors also play a role: Cheng et al. (2024) proposed a dynamic adjustment method that uses risk attitudes to update preference information across changing decision contexts. They have developed a risk attitudes identification method and creatively combined it with the DEMATEL technique (Taherdoost and Madanchian, 2023) to group decision makers into four categories including core decision makers, driven decision makers, independent decision makers, and followers.

Other MCDM research has focused on consistency and tolerance for inconsistency. Korhonen et al. (2012) found significant inconsistency in value-function use across individuals. When utilizing any particular value function in binary choices, individuals exhibit a lack of consistency. A significant portion of participants achieved consistency after eliminating 10% of responses.

Building upon these insights, Dynamic Multi-Criteria Decision Making (DMCDM) has emerged as a valuable paradigm for modelling preferences that evolve over time (Benítez et al., 2020; Yang et al., 2025). Unlike traditional static MCDM, DMCDM explicitly incorporates temporal information, aggregating evaluations across multiple stages. However, current DMCDM approaches largely focus on past and present information while neglecting the evolving nature of future preferences (Yang et al., 2025).

The study performed in this paper incorporates some DMCDM ideas in the DEX context, specifically addressing the evolving nature of DEX decision tables. Positioned within this dynamic MCDM landscape, our study provides one of the first empirical examinations of how preference knowledge expressed through DEX decision rules and weights varies across individuals and time.

3. Methods

To answer the research questions, we created a questionnaire that was applied two times: at the end of November 2023 and in the middle of January 2024 (48 days apart). The study involved 109 students in the fields of informatics and entrepreneurship. The time difference of 48 days was chosen according to the course schedules, so that the survey was applied at the beginning and end of the semester. This was essential in order to address the same participants in a well-controlled time frame, which was missing in the preliminary study. While focusing on a single group, i.e., students, may be considered a limitation, it enabled the development of a questionnaire tailored to decision problems relevant to that group. This would not have been feasible with a more heterogeneous group and would likely have reduced participants' motivation.

The questionnaire consisted of the following parts:

1. Data about the respondents included the password, gender, birth year, average grade achieved in the study, the type of study (undergraduate or graduate), the year of study, and the field of study. To achieve data privacy, in the first data-collecting round, the respondents had to choose the password they wanted and remember it until the second round. Consequently, we used the password to match answers from the same students in the two rounds.
2. Data forms related to the four use case studies: study program, student success, car, and store. The case studies are further explained below. Data forms related to each case study included (A) the questions related to the quantified distribution of the case study's criteria weights, and (B) the definition of decision rules.
3. The final section related to the conditions in which the data-collecting procedure was implemented. The questions from this section were related to the clarity of the questionnaire, the appropriateness of the number of questions in the questionnaire, and the presence of interfering factors during the data collection (noise, lack of time, misunderstanding the questions, the influence of colleagues/friends, lack of concentration, and lack of interest in the topic).

In the step 2 above, the participants were requested to define decision rules in four predefined decision tables. The response time was not limited (but was typically well within the 20-minutes range). Since all participants were from Croatia, the questionnaire was formulated in Croatian; an English translation is presented hereafter. The questionnaire was implemented in an online environment, using a Google Sheets document. The survey was anonymous; however, to be able to connect responses at different times, participants defined their passwords and used the same password both times.

3.1. Use cases

Use case 1: Study

The first case study was related to the selection of the *Study* program, and was included in this experiment as a topic of general interest for the participating students. There were three predefined criteria for this case study:

- *Match* of the study program *with personal interests* and preferences: "low", "medium" or "high";
- *Employability* after graduation: "low", "medium" or "high";
- The *complexity* of the study program for the student: "high", "medium" or "low".

There were four possible outcomes in this case study: "unacc", "accept", "good", and "excel". There were 27 possible combinations of the three criteria values, since each could achieve three possible values. An empty decision table for this case study is presented in Table 1. The participants were requested to mark exactly one choice in each row under *Study program*.

Further, quantitative weights of criteria were collected using the form presented in Table 2. Before providing their weight assessments, participants received the following instructions: "Weights represent the relative importance of each criterion in your overall decision. A higher weight means that the criterion contributes more strongly to the final evaluation. The total must sum to 100." We additionally explained that weights can be interpreted as the percentage contribution of each criterion to the overall decision.

Use case 2: Success

The second case study was related to the expected *Success* of students during their studies. Here we defined three criteria that were further evaluated by students. They are:

- Intrinsic *motivation* for studying: "low", "medium" or "high";
- *Organizational skills* (time and resource management): "low", "medium" or "high";
- Critical thinking and *problem-solving skills*: "low", "medium" or "high".

There were five possible outcomes for this case study: "very low", "low", "medium", "good", and "excellent". Here again, there are 27 possible value combinations and the data collection forms were similar to Tables 1 and 2.

Use case 3: Car

The third task (*Car*) was retained from the preliminary study (Bohanec, 2023). The task is to define decision rules for evaluating a family car considering just two criteria:

- *Price*: "high", "medium" or "low";
- *Technical characteristics*: "poor", "acceptable", "good", and "excellent".

There are four possible outcomes: "unacc", "accept", "good", and "excel".

An empty decision table consisting of 12 possible combinations of the criteria's discrete values was presented to the respondents, asking them to mark the corresponding values of *Car* (Table 3). In connection with this table, respondents were also asked to assess the weights of the two criteria, as shown in Table 4.

Use case 4: Store

The fourth case study (*Store*) was also retained from Bohanec (2023). It was originally inspired by the experiment designed by Vetschera et al. (2014), but uses a reduced number of categories to keep the decision table reasonably small. The task is to assess the suitability/attractiveness of the store for daily purchases, primarily referring to purchases of groceries. There are four qualitative criteria:

- *Store size*: "market" or "supermarket";
- *Walking distance from home:* "less than 10 minutes", "more than 10 minutes";
- *Price category*: "lower", "higher";
- Product *quality*: "lower", "higher".

There are the same four possible outcomes as with *Car*: "unacc",

Table 2
Question to assess the study program weights.

Please assess criteria weights so that their sum equals 100:

Criterion	Weight
Match of the study program with personal interests and preferences	
Employability after graduation	
The complexity of the study program for me	
Sum	**100**

Table 3
A table for acquiring decision rules for the evaluation of cars.

	Price	Tech.char.	Car			
			unacc	accept	good	*excel*
1	high	poor				
2	high	accept				
3	high	good				
4	high	*excel*				
5	medium	poor				
6	medium	accept				
7	medium	good				
8	medium	*excel*				
9	*low*	poor				
10	*low*	accept				
11	*low*	good				
12	*low*	*excel*				

Table 4
Questions to assess car weights.

Please assess criteria weights so that their sum equals 100:

Criterion	Weight
Price	
Tech.char.	
Sum	**100**

"accept", "good", and "excel".

Notice that the four *Store* criteria are binary (two-valued); this yields 16 possible value combinations, which are presented in the questionnaire in a table similar to Table 3. An analogous question to that from Table 4 is asked for the *Store* task, too.

The fourth task is considered more difficult than the other ones. Even though there are fewer decision rules, the task entails the simultaneous consideration of four criteria, which is typically more complex than combining only two or three. The questionnaire itself does not interpret the criteria and their values any further, so we may expect subjective individual interpretations. In contrast with previous use cases, where all the criteria are clearly preferentially ordered and we may expect that decision rules will reflect this order, this is much less so with *Store*. Buying habits largely differ between consumers, and while one may prefer a "lower" price category, some other may equally well prefer the "higher". Thus, we can hardly expect any clear preferential ordering of rules in the *Store* case.

3.2. Differences between decision tables

Research questions A and B require the calculation of differences between two decision tables. Given a decision table template, such as Table 3, a participant marks exactly one choice among the possible outcomes in each of the rows. For brevity, instead of using words, such as "unacc", "accept", "good" and "excel" for outcomes in Table 3, we represent marks by their ordinal numbers, in this case 1 to 4. Since Table 3 consists of 12 rows, the marks of some respondent form a vector consisting of 12 ordinal numbers, for example $\langle 111112231234 \rangle$.

Generally, a table T_t can be thus represented as:

$$T_t = \left\langle y_{t,1}, y_{t,2}, \ \ldots, y_{t,k} \right\rangle$$

This is a vector of k ordinal numbers $y_{t,r} \in \{1, 2, \ldots, m\}, r = 1, \ldots, k$, where r is the rule index, m is the number of output values, and k equals to the size of the decision table (number of decision rules). For the four use cases, $m = 4, 5, 4, 4$ and $k = 27, 27, 12, 16$, respectively. Notice that the lowest and highest possible vectors are $\underbrace{\langle 111\ldots1 \rangle}_{k}$ and $\underbrace{\langle mmm\ldots m \rangle}_{k}$. This gives the following formula for calculating the difference between two decision tables:

$$\Delta T_a, T_b = \frac{1}{k} \sum_{i=1}^{k} \frac{|y_{a,i} - y_{b,i}|}{m - 1}$$

This formula is designed so that it yields the difference of 1 for the above extreme case, and 0 for two equal decision tables.

3.3. Consistency of decision rules

Whenever value scales of all involved criteria are preferentially ordered (from 'bad' to 'good' values or vice versa), we can assume that "rational" decision rules will be *consistent*: the better the input criteria (such as *Price* and *Technical characteristics*), the better the outcome (*Car*). Consequently, the decision table is expected to obey the *principle of dominance* (Kourouxous and Bauer, 2019), so that the aggregation function is *monotone*: in the direction of each improving criterion, the outcome improves or stays constant at least.

To detect monotonicity violations, we compared each decision rule with all rules that dominate it in the decision table. A rule *A dominates* rule *B* if *A* is equal or better in all criteria and strictly better on at least one criterion. In this case, *B* is considered *dominated* by *A*. A violation occurs if the dominated rule produces a strictly better output value than the dominating rule.

For each participant we calculated the proportion of non-violated (monotone) rule pairs. In principle, well-defined decision tables are expected to be consistent, except for the *Store* use case, which is expected to involve non-monotone concepts and was included in the study to assess the average level of (in)consistency in such cases.

3.4. Assessment of weights from decision tables

Even though DEX is a qualitative method, for which the concept of criteria weights is somewhat unnatural, it is possible to approximately assess weights from a defined decision table. In this study, we used three methods. The first two, *Gini gain* (WGG) and *Information gain* (WIG), are routinely used in machine learning for determining the strength of features from data (Raileanu and Stoffel, 2004; Rokach and Maimon, 2015, pp. 62–63). These methods employ *impurity measures Gini* and *Entropy*, respectively, to assess the disorder of data, represented by a DEX decision table T:

$$Gini(T) = 1 - \sum_{i=1}^{n} p^2(v_i)$$

$$Entropy(T) = -\sum_{i=1}^{n} p(v_i)\log_2 p(v_i)$$

Here, $p(v_i)$ is the probability/proportion of value v_i occurring in the table and n is the number of input criteria ($n = 3,\ 3,\ 2,\ 4$ for the four use cases, respectively).

Using these measures, a relative weight RW of some criterion c is determined as

$$RW(c,T) = M(T) - \sum_{v \in S_c} \frac{|T_{c=v}|}{|T|} M(T_{c=v})$$

where M is an impurity measure (*Entropy* for WIG and *Gini* for WGG), v are values taken from S_c, the qualitative value scale of c, and $T_{c=v}$ denotes the part of T where $c = v$. $|T|$ denotes the size of table T in terms of the number of decision rules.

The third method, *Linear approximation* (WLA), is implemented in DEX software (Bohanec, 2025). It interprets decision rules as points in a multi-dimensional space and approximates them with a hyperplane using the least squares principle. Criteria weights are approximated from the slopes of the hyperplane: the higher the slope in the direction of a criterion, the higher the corresponding relative weight of the criterion. For more details about WLA, see Bohanec and Zupan (2004, sec. 3.4) and Deguine et al. (2021, sup. sec. 2).

Among the three methods, WLA is better suited for consistent, monotone and potentially linear decision tables, such as *Car*, while WGG and WIG might better capture the importance of variables in non-monotone data tables, such as *Store*.

3.5. Comparison of weights

In order to compare weights $W_M = w_1, \ldots, w_n$, estimated by some method M from decision rules (hereafter called “quantified weights”), and weights $\Omega = \omega_1, \ldots, \omega_n$ as given by the respondent (“quantitative weights”), we used the formula:

$$\Delta W_M, \Omega = \frac{1}{n} \sum_{i=1}^{n} \frac{|w_i - \omega_i|}{100}$$

Here, n represents the number of criteria. The range of weights w_i and ω_i is $[0, 100]$. Then, the range of $\Delta W_M, \Omega$ is from 0 (no difference) to 1 (extreme difference). The same formula is applicable for comparing quantified weights obtained by any two weight-assessment methods.

In this study, we made several types of weight comparisons. The first was to assess the *match* between quantitative and quantified weights obtained in by the same respondent in a single questionnaire, using the above formula for different methods M. In this way we assessed the differences caused by using different methods: direct formulation of quantitative weights and indirect formulation through decision tables. The second comparison type involved weights collected at two different time points, which was aimed at assessing the stability of formulations.

4. Results

In this section, we present the research results. The questionnaire surveys were carried out on November 28th, 2023 and January 15th, 2024 (48 days difference).

4.1. Collected data

The data collecting procedure was organized at the University of Zagreb, Faculty of Organization and Informatics, Varaždin, Croatia, as a regular assignment in seminar classes on two courses: Business decision analysis (graduate study “Economics of Entrepreneurship”, 1st year) and Design thinking in digital transformation (undergraduate study “Information Systems”, 3rd year). There were 111 and 112 questionnaires completed in the two surveys, respectively (223 in total). Among these, 109 pairs were identified to belong to the same participant.

Table 5 shows the number of students and their average grades (in brackets, expressed on the scale 1.00-5.00) with respect to gender and the study program.

4.2. Differences between decision rules

Inter-personal differences between decision rules (research question A1) were assessed on 223 questionnaires collected in both surveys. Table 6 shows the minimal and maximal vectors observed in the experiment, and distances between them. The results indicate that respondents’ answers cover a large proportion of decision space between the possible extreme outcomes, and that the maximum distances between vectors are large.

Observing individual vectors also reveals a great variability of answers (Table 7), as most of the respondents’ vectors are distinct. Even with *Car*, which is characterized by a smaller decision table and less possible input value combinations than other use cases, there were as many as 156 distinct answers. There was a single most frequent vector ⟨112312341234⟩, which was given 16 times. In other use cases, almost

Table 5
Research sample: the number of participants and their average grades.

	Undergraduate study (Informatics)	Graduate study (Economics)	Total
Male	24 (3.13)	8 (2.92)	32 (3.07)
Female	24 (3.17)	53 (3.46)	77 (3.37)
Total	48 (3.15)	61 (3.39)	109 (3.28)

Table 7
Most frequent vectors and the number of distinct vectors.

Use case	Most frequent vector	Occurred [times]	Distinct vectors [of 223]
Study	⟨111111111111222333111333444⟩	3	221
Success	N/A (all vectors were distinct)	1	223
Car	⟨112312341234⟩	16	156
Store	⟨1413141314131413⟩	4	216

Table 6
Minimal and maximal observed vectors of respondents’ answers in decision tables.

Use case	Min. vector	Max. vector	Distance
Study	⟨111111111111111111111111111⟩	⟨334334444334444444444444444⟩	0.926
Success	⟨111111111111112122111112223⟩	⟨233354445445555555545555555⟩	0.787
Car	⟨111111211111⟩	⟨244434444444⟩	0.889
Store	⟨1111111112111111⟩	⟨4434444444444444⟩	0.958

Table 8
Average distances between vectors.

Use case	*Inter-personal*	*Intra-personal*	*Ratio Intra/Inter*
Study	0.214 ± 0.077	0.154 ± 0.076	0.720
Success	0.167 ± 0.059	0.131 ± 0.055	0.784
Car	0.157 ± 0.078	0.122 ± 0.074	0.777
Store	0.235 ± 0.085	0.180 ± 0.088	0.766

Table 9
Average consistency of decision rules.

Use case	Average consistency	Fully consistent tables [%]
Study	0.970	32.3
Success	0.986	46.6
Car	0.982	71.3
Store	0.768	4.0

all vectors were distinct; only 3 and 4 repetitions were observed in *Study* and *Store*, respectively. In *Success*, all vectors were distinct.

Average distances between these vectors are shown in the *Inter-personal* column of Table 8. They were calculated on all pairs of vectors collected in both surveys (24753 pairs). The differences are roughly between 16% (*Car* and *Success*) and 24% (*Store*).

Intra-personal differences between decision rules (research question B1) were assessed on 109 pairs of questionnaires that were answered by the same participants at two different times. Average distances between vectors in this case are shown in the third column of Table 8. They are between about 12% (*Car*) and 18% (*Store*).

As expected, intra-personal differences are smaller than inter-personal ones. However, they are not *substantially* smaller: the ratios between inter- and intra-personal differences (Table 8) are all in the 0.7–0.8 range and are surprisingly similar to each other. To put it

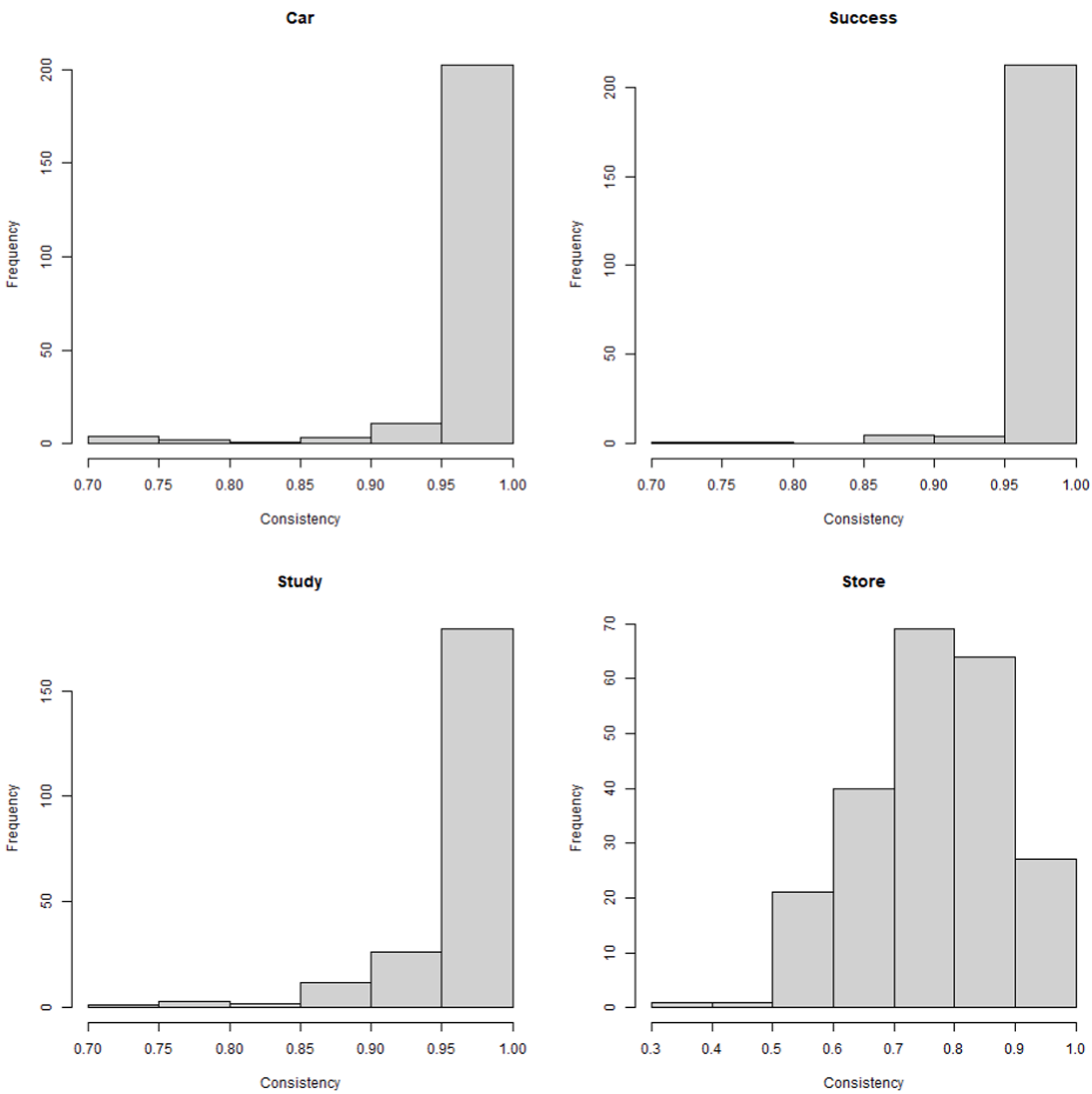

Fig. 1. Consistency distributions in the four use cases.

loosely: the variation in a person's preferences between the two points in time is about three-quarters of the variation between different individuals.

These results indicate that subjects' preferences drift a lot over time and that it is difficult for individuals to provide the same decision rules twice.

At this point, this study cannot really explain the drift; we can only speculate about the effects of changed preferences, changed decision context, bad memory and even imprecise, inaccurate or otherwise "elusive" nature of decision rules.

4.3. Consistency of decision rules

The average consistency (research question C1) of decision rules (Table 9) was high (greater than 97%) with use cases *Study, Success* and *Car*, and substantially lower (77%) with *Store*. However, the proportion of participants that were able to construct fully consistent tables was not as high, spanning between 71% (*Car*) and 32% (*Study*), with a notable and expected exception of *Store* (4%).

Fig. 1 compares consistency distributions in the four use cases. The consistencies of *Car* and *Success* decision rules are indeed high and mostly close to 1.0. The consistencies of *Study* are generally worse, which may indicate that the decision problem is more difficult than *Success* and/or involves non-monotone concepts. *Store* is an expected outlier, indicating that the involved concepts are largely non-monotone.

To what extent do consistencies of decision rules acquired from the same subject at different times differ (research question C2)? Table 10 presents the change in terms of the number of decision tables whose consistency decreased (label "Lower"), increased ("Higher") or remained the same ("Equal") over time. The results indicate that all three outcomes did occur, but in different shares depending on the use case. In use cases *Study* and *Success* most of the students increased their consistency. In the case of *Car*, most students achieved the same consistency in both rounds. Finally, in the case *Store*, most students decreased their consistency.

Table 10
The change of decision rules consistency over time.

	Lower	Equal	Higher
Study	38	21	50
Success	40	28	41
Car	15	70	24
Store	53	7	49
Total	146	126	164

4.4. Assessment of weights

The remaining set of research questions is concerned with criteria weights. All the participants were able to completely define their subjective criteria weights so that their sum equals 100, as required in the questionnaire (see examples in Tables 2 and 4). Consequently, acquiring criteria weights directly from participants seems an easy task. However, we are further interested in the "correctness" of the provided answers, their alignment with quantified weights from decision tables, and how they differ among participants and over time.

First, how much do *quantified* weights (those inferred from decision rules) and *quantitative* weights (as provided by the participants) differ (research question D)? Fig. 2 displays a boxplot of weights of *Car* decision rules as assessed by the participant (Ω) and by the three methods defined in Section 3.4: WLA, WIG and WGG. On the one hand, we can see that participants, in average, assessed the two criteria, *Price* and *Technical characteristics*, almost equally, with a slight statistical leaning towards the latter. On the other hand, all the quantification methods clearly indicated that *Technical characteristics* are far more important than *Price*. Here we can claim that users did not assess their weights really well, while the methods were largely consistent with each other.

Results for *Store* are shown in Fig. 3, where participants' weights (Ω) are compared with those assessed by the WIG method. Again, comparing human and algorithmic weights, the former are less extreme and all lean towards 20–30%, while the latter are more extreme, ranging from about 10% to 50%. Yet again, the three algorithmic methods turned out similar to each other (these results are not shown here). In contrast with *Car*, the order of criteria's importance was estimated almost correctly by the participants (*Size* being the least, and *Price* and *Quality* the most important). However, there is a striking difference between the participants' assessment of *Price* and *Quality* (almost equal around 30%) and the WIG's, which indicates a large difference (20% vs. 50%).

In contrast with *Car* and *Store*, the results for *Study* and *Success* (Fig. 3) indicate a good match between the participants' and algorithms' assessment of weights.

The results of comparing all pairs of weight assessments (by Ω, WLA, WIG and WGG) and calculating their difference using the formula from Section 3.5, are shown in Table 11. Particularly small differences (in the

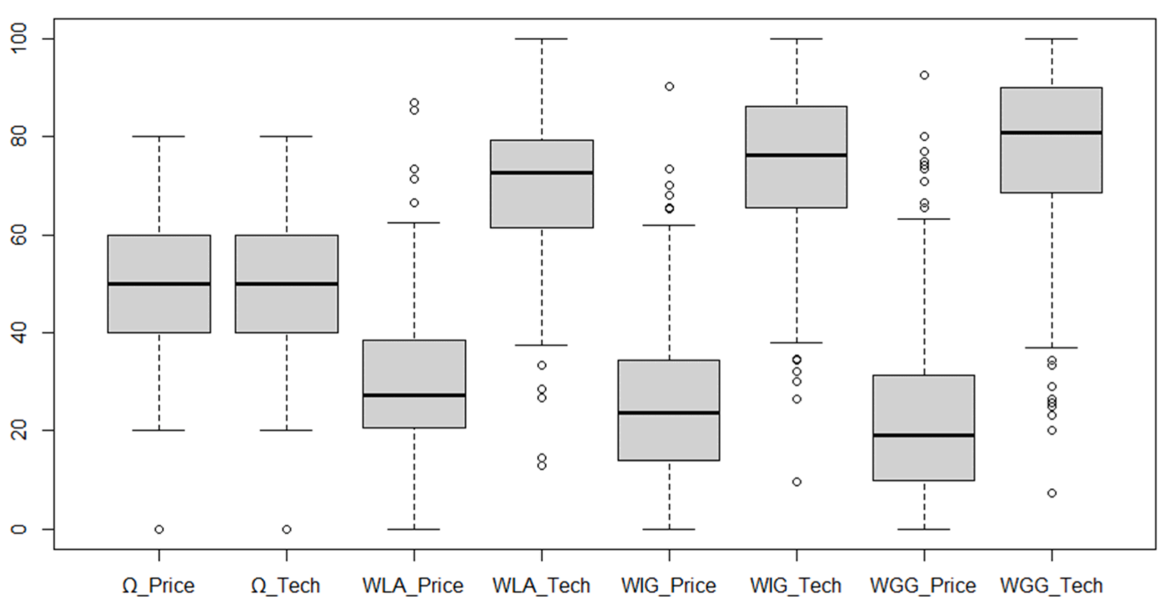

Fig. 2. Weights of Car criteria assessed by the participant (Ω) and three methods: WLA, WIG and WGG.

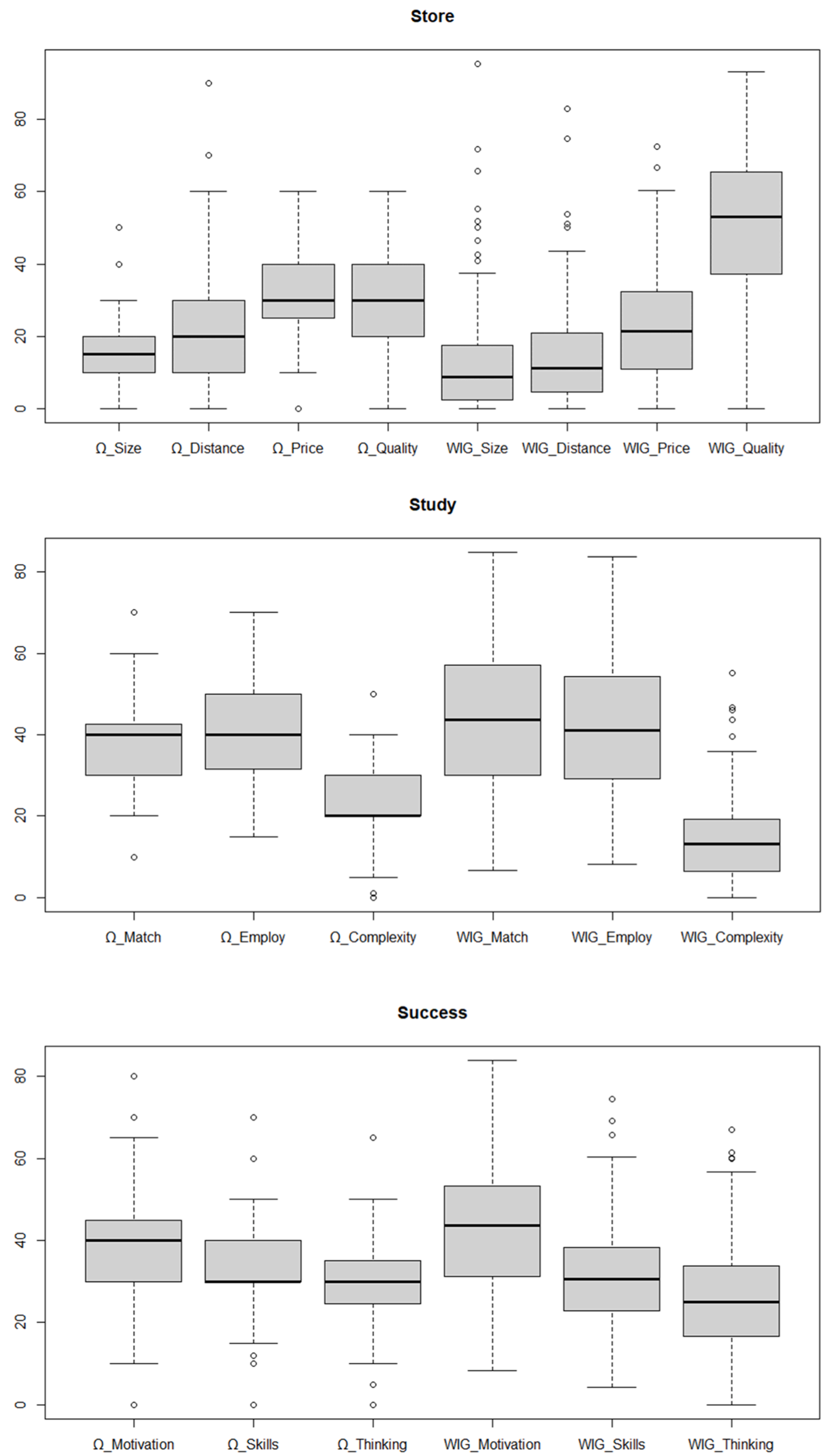

Fig. 3. Weights of store, study and success criteria assessed by the participant (Ω) and method WIG.

Table 11
Differences between weight assessments of participants (Ω) and methods WLA, WIG and WGG.

	Study			*Success*			*Car*			*Store*		
	WLA	WIG	WGG	WLA	WIG	WGG	WLA	WIG	WGG	WLA	WIG	WGG
Ω	0.108	0.127	0.148	0.098	0.115	0.131	0.229	0.260	0.292	0.162	0.157	0.178
WLA		0.058	0.076		0.048	0.069		0.077	0.096		0.068	0.075
WIG			0.039			0.036			0.038			0.044

Table 12
Differences among quantitative weights, formulated by participants, with respect to each case study and time. Labels C1–C4 refer to individual use-case criteria.

		Round 1 (November 2023)		Round 2 (January 2024)	
		Range	Mean±stdev	Range	Mean±stdev
Study	C1	10–70	36.35±11.19	20–70	37.48±10.92
	C2	20–70	43.19±9.76	15–70	39.89±10.96
	C3	0–50	20.46±9.18	0–50	22.00±8.25
Success	C1	10–80	38.98±12.95	15–70	36.07±10.33
	C2	0–60	32.42±10.37	12–70	33.75±8.84
	C3	5–50	28.32±10.06	0–65	30.20±8.85
Car	C1	0–70	47.92±12.19	30–80	51.28±10.51
	C2	0–80	50.24±12.37	20–70	48.72±10.51
Store	C1	0–50	15.50±9.45	0–40	17.52±9.61
	C2	0–50	19.61±12.33	0–90	22.61±14.8
	C3	10–60	33.31±10.14	0–60	28.99±10.69
	C4	0–60	31.22±10.09	0–60	29.17±11.07

4% range) are between WIG and WGG, which is not surprising considering the similarity of methods. Differences between WLA vs. WIG and WGG are in the range from 5% to 10%. On the other hand, differences between Ω and the remaining three methods are all greater than 9% and reach almost 30% with *Car*. This supports observations from Figs. 2 to 3 that indicate large discrepancies between humans' and algorithm-assessed weights. However, coming as a surprise and contrary to our expectations, the differences between Ω and WLA turned out substantially lower than those of Ω vs. WIG and WGG (except by a small margin with *Store*). This indicates that the method WLA, as implemented in DEX software, actually resembles user-assessed weights relatively well, even for non-monotone decision tables.

How much do quantitative criteria weights acquired from different subjects differ (research question A2)? Table 12 presents the differences among quantitative weights per each case study and data collection time. Although the mean weights of different criteria vary significantly, the standard deviation is stable (from 8.25 to 12.92).

Table 13 compares weights between the first and second surveys for the 109 participants that answered the questionnaire twice (research question B2). The most important observations are that weights do change in time and the participants' assessments (Ω) change less than those of the three methods. Participants' weights changed about 9% between the two surveys, while algorithm-assessed weights changed between 10% and 27%. Among the latter, the largest changes were observed in the *Car* use case (all in the 15% to 27% range). In the remaining three use cases, changes were roughly between 10% and 19%.

The differences among criteria weights acquired from the same subject at different times (research question B2) are presented in Table 14, separately for quantitative and quantified criteria. The highest differences are achieved in the study *Store* in the case of quantitative criteria and in the study *Car* in the case of quantified criteria. Generally, average distances between quantified weights are almost twice as large as between quantitative weights. This corroborates (and quantifies) that algorithm-assessed weights are more pronounced than those provided by the participants.

5. Discussion

This study explores the variability and consistency of decision-making preferences, expressed through decision rules and criteria weights, using the qualitative DEX method. The findings highlight key aspects of human decision-making, including the variability and (im) precision of weights in multi-criteria decision-making (MCDM).

One of the central findings of this study is the significant variability in decision-making preferences both across individuals (inter-personal) and within the same individual over time (intra-personal). The intra-personal variability, accounting for approximately 75% of inter-personal differences, challenges the assumption of stable preferences often embedded in MCDM methodologies. This variability raises questions about the reliability of models that assume static weights or decision rules. Real-world decision-making often involves evolving priorities and contextual influences, suggesting that MCDM methods might benefit from dynamic or adaptive approaches that accommodate such changes.

The study reveals that while consistency in decision rules is relatively high in cases with clear preferential order (e.g., use-cases *Car* and *Success*), the stability of weights and rules remains low. This dichotomy suggests that while participants can adhere to logical consistency within a decision model, their preferences are less stable over time. For MCDM designers, this highlights the importance of tools that ensure consistency while acknowledging and adapting to preference drift. Automated consistency-checking tools, for instance, could support users in maintaining logical coherence in decision rules while allowing flexibility in adapting weights over time.

Table 14
The differences among quantitative and among quantified criteria weights.

	$\Delta W_M\Omega$ (quantitative)				$\Delta W_M\Omega$ (quantified)			
	Min	Max	Mean	St.dev	Min	Max	Mean	St.dev
Study	0	0.33	0.084	0.063	0	0.38	0.155	0.087
Success	0	0.40	0.086	0.064	0	0.38	0.124	0.074
Car	0	0.50	0.080	0.095	0	0.77	0.158	0.142
Store	0	0.45	0.094	0.065	0	0.48	0.136	0.078

Table 13
Weight differences between the two surveys.

	Study				*Success*				*Car*				*Store*			
	Ω	WLA	WIG	WGG	Ω	WLA	WIG	WGG	Ω	WLA	WIG	WGG	Ω	WLA	WIG	WGG
Ω	0.084	0.111	0.132	0.154	0.089	0.110	0.128	0.145	0.080	0.204	0.234	0.266	0.095	0.170	0.165	0.185
WLA		0.110	0.127	0.139		0.097	0.117	0.133		0.141	0.163	0.182		0.129	0.134	0.143
WIG			0.133	0.142			0.124	0.138			0.158	0.165			0.136	0.148
WGG				0.152				0.153				0.168				0.146

The quest for precise relative weights appears less justified given the observed variability. Quantitative weights assessed by participants showed a 9% drift over 48 days, while algorithmically derived weights varied between 10% and 27%. These findings suggest that precise weights may not reliably capture the decision-maker's preferences over time, potentially leading to misplaced confidence in their stability. Instead, an argument can be made for prioritizing "approximate" weights that reflect general preferences and are robust to variability.

The discrepancies between algorithmically derived weights and human-assessed weights raise intriguing questions about the role of algorithms in decision support. While algorithms such as WIG, WGG, and WLA consistently identified certain criteria as more influential, participants often assigned more balanced weights. This divergence underscores a potential tension between human intuition and algorithmic precision. Designing systems that bridge this gap—by enhancing algorithm transparency or aligning algorithms more closely with user expectations—could improve trust and usability in MCDM tools.

Our research, along with previous studies, demonstrates that subjects' preferences are susceptible to situational influences and exhibit drift over time. This means that the judgments decision-makers assign to different outcomes are not fixed but can be altered by factors such as the decision-maker's bias, time pressure, problem complexity, decision-maker's level of concentration, etc. The open-ended question in this study asked respondents to identify factors that influenced their preferences. The results showed that 3% cited time constraints, 7% mentioned environmental noise, 3% referred to the influence of colleagues, 29% indicated the possibility of losing concentration, and 22% suggested that a lack of interest in the topic may have affected their preferences.

The results of our research also align with the KoW paradigm (Cash and Oppenheimer, 2025) by demonstrating that participants' preferences are sensitive to situational factors and exhibit instability due to influences such as deconcentration and lack of interest. The KoW paradigm offers an essential framework for examining how decision-makers (mis)interpret these situational influences, which, as our findings confirm, directly contribute to variability and instability in both perceived criteria weights and final decisions.

Our findings, supported by previous research presented in this work, can have significant implications for the design and implementation of MCDM systems. MCDM models must be dynamic, incorporating mechanisms to adapt to evolving preferences. This includes periodic recalibration of weights or decision rules. A promising approach to addressing these issues is the application of Dynamic Multi-Criteria Decision Making (DMCDM). In DMCDM problems, the impact of decision information from different periods on the final decision outcome varies significantly due to the timeliness of the information. Accurately determining the weights of different periods is crucial for aggregating this information effectively. This process necessitates a careful consideration of both subjective factors, such as the preferences of decision-makers, and objective factors derived from the decision matrices. While previous research has explored various approaches for determining period weights, the work of Yang et al. (2025) represents a significant advancement. They developed an optimization model that effectively integrates subjective preferences of decision-makers with objective information extracted from the decision matrices to determine period weights. To predict future alternative performance, machine learning methods were employed to analyze historical data, including the identification of (non)linear trends. Benítez et al., (2020) applied the AHP method within an objective framework, where weight assignments are stochastically calculated instead of being defined based on expert judgment. The main objective of their study is to propose a dynamic decision model based on AHP for the maintenance planning of reinforced concrete structures under corrosion risk. This approach also exemplifies the application of DMCDM. Campanella and Ribeiro (2011) propose a flexible framework for DMCDM, extending the established theoretical foundation. The framework's applicability to a diverse range of dynamic decision-making problems is illustrated through a case study involving helicopter landings.

There is a limited number of such approaches available, and their development is still ongoing. Previous models have primarily focused on historical and current data, neglecting future prediction and adaptation to change. Therefore, developing DMCDM models that can predict future alternative performance is a significant challenge for future research.

Considering the implications for further work, we emphasize the importance of decision makers' training. Training modules can enhance decision-making by helping users develop more stable and consistent preferences, particularly in high-stakes applications. By mitigating the impact of factors such as time pressure, distractions, and the complexity of the problem, training can improve decision quality. Furthermore, ensuring that decision makers understand how algorithms derive weights and decision rules is crucial for bridging the gap between algorithmic insights and human judgment.

6. Conclusion

The main aim of this study was to assess the stability and consistency of decision tables and weights in the method DEX, considering both the differences between different people and differences between the same individuals in different points in time. Our assumption when conceiving this study was that the DEX decision tables were somewhat "rigid", as they consist of just a few discrete input-value combinations (rules) and a few discrete output values that can be assigned to each rule. Consequently, it might seem that there is little freedom in defining decision rules and that all decision tables might look the same. Both the preliminary and current study indicated just the opposite. It turned out that people – and their decision preferences – are incredibly diverse.

The main contributions of this study are the identification and quantification of the variability and consistency of DEX decision rules. Almost all decision tables, obtained from participants in this experiment, were distinct. Furthermore, people change in time, and their preferences, expressed in terms of decision rules and criteria weights, are far from stable.

The most important findings of this study are:

- *Inter-personal differences* (decision makers' preferences, expressed by different participants in terms of decision tables) are very diverse. The maximum difference between rule vectors, measured in the [0, 1] range, varied between 0.79 (use case *Success*) and 0.96 (*Store*). The average differences were also high, between 16% (*Car* and *Success*) and 24% (*Store*).
- *Intra-personal differences* of decision rules of a single decision maker also change in time. Not as much as inter-personal differences, but close (between 12% and 18% in this study).
- Intra-personal differences are not substantially smaller than inter-personal ones. In the 48 days, preferences of individual participants changed by about three quarters of the differences observed between different participants.
- The *consistency* of decision tables was high (97–99%) in all use cases involving preferentially ordered criteria (i.e., all but *Store*). This indicates that the majority of participants were able to adhere to the principle of dominance without any hints or supporting tools.
- On the other hand, the percentage of *fully consistent* decision tables was not as high, ranging from 33% to 71%. This indicates that defining fully consistent decision tables is still hard and that some form of algorithmic consistency checking would be beneficial.
- The *weights of criteria*, which are provided by the user (quantitative weights) and assessed from decision rules (quantified weights), also vary in many ways. Considering all criteria in a given use-case context, it is notable that the differences between quantitative weights are much smaller than between the quantified ones; the average distances between the latter are almost twice as large as

between the former. Quantitative weights also tend to change less in time (about 9% in 48 days) than the quantified ones (10–27%).

Overall, this study indicates that human decision-making preferences are very diverse and volatile. Any MCDM method, including DEX, should consider that preferences can span over the whole "decision space", whatever it is (decision tables in DEX). The understanding and interpretation of the same concepts, such as criteria weights, may vary between the people and algorithms, and also between different algorithms (WIG, WGG and WLA generally yield different weights on the same data). Any acquired preference information can change in time. In this light, the quest for obtaining as "precise" criteria weights as possible, which is pursued in many MCDM methods, may be somewhat relaxed in the context of dynamic MCDM. In addition to providing a comprehensible and safe environment for acquiring and representing decision preferences, MCDM methods should take more care of inevitable inter- and intra-personal differences while supporting the process with mechanisms that can be automated, such as consistency checking.

All these findings challenge traditional decision theories that assume stable preferences and suggest that real-world decision-making is far more dynamic and variable. Models like prospect theory or bounded rationality might be enriched by incorporating variability and consistency as core components of decision-making processes. Moreover, the study underscores the value of qualitative approaches like decision rules in capturing the complexities of human preferences, particularly in contexts where numerical weights might not fully capture subjective priorities.

The findings, reinforced by prior research, bring important implications for the design and implementation of future MCDM systems:

- *Dynamic Adaptation*: MCDM models could incorporate mechanisms to adapt to preference drift, such as periodic recalibration of weights or decision rules.
- *User Training*: Training modules could help users develop more stable and consistent decision preferences, particularly in high-stakes applications.
- *Transparent Algorithms*: Ensuring that users understand how algorithms derive weights and decision rules could bridge the gap between algorithmic insights and human judgment.

This study represents an initial step in this research area, with limitations related to sample size and the analysis of preference changes. The study was limited to student participants from a particular university, which restricts the generalizability of the findings to other populations, for example, professionals or older adults. While we may expect similar behavioral patterns there, this has to be confirmed by further studies. Also, this study was restricted to a one-time interval of 48 days. A relevant question is how the observed characteristics change over different periods of time. This requires longitudinal studies, which are difficult to carry out, because we need to ask the same participants the same questions multiple times, while affecting their preferences as little as possible. Therefore, future research should focus on conducting more in-depth analyses of preference dynamics, utilizing more representative samples, developing innovative DMCDM approaches and exploring the impact of using approximate versus precise weights on decision outcomes, particularly in dynamic contexts.

Ethical considerations

The data for this study was collected on a voluntary basis using anonymous questionnaires. No personal data of participants was collected. According to the opinion of the Ethical Committee of the Faculty of Organization and Management of the University of Zagreb, Ref. 2186-62-06-20-3 and 2186-62-14-23-21, this research is ethically acceptable and is in accordance with the Code of Ethics of the University of Zagreb.

CRediT authorship contribution statement

Marko Bohanec: Writing – review & editing, Writing – original draft, Supervision, Methodology, Formal analysis, Conceptualization. **Nikola Kadoić:** Writing – review & editing, Writing – original draft, Investigation, Formal analysis, Conceptualization. **Nina Begičević Ređep:** Writing – review & editing, Writing – original draft, Investigation, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Marko Bohanec reports financial support was provided by Slovenian Research and Innovation Agency. Nikola Kadoic reports financial support was provided by Croatian Science Foundation. Nina Begicevic Redjep reports financial support was provided by Croatian Science Foundation.

Acknowledgements

The first author acknowledges the financial support from the Slovenian Research and Innovation Agency for the programme Knowledge Technologies (research core funding No. P2-0103). The second and third authors acknowledge the support from the Croatian Science Foundation under the project HRZZ-IP-2020-02-5071 HELA (Improve HEI Maturity to Implement Learning Analytics).

References

Ariely, D., Loewenstein, G., 2000. When does duration matter in judgment and decision making? J. Exp. Psychol. Gen. 129 (4), 508–523. https://doi.org/10.1037/0096-3445.129.4.508.

Benítez, P., Rocha, E., Varum, H., Rodrigues, F., 2020. A dynamic multi-criteria decision-making model for the maintenance planning of reinforced concrete structures. J. Build. Eng. 27. https://doi.org/10.1016/j.jobe.2019.100971.

Bohanec, M., Zupan, B., 2004. A function-decomposition method for development of hierarchical multi-attribute decision models. Decis. Support Syst. 36, 215–233. https://doi.org/10.1016/S0167-9236(02)00148-3.

Bohanec, M., 2022. DEX (Decision EXpert): a qualitative hierarchical multi-criteria method. In: Kulkarni, A.J. (Ed.), Multiple Criteria Decision Making, Multiple Criteria Decision Making, 407. Springer, Singapore, pp. 39–78. https://doi.org/10.1007/978-981-16-7414-3_3.

Bohanec, M., 2023. Inter- and intra-personal differences, and consistency of decision rules, in Multi-criteria modelling method DEX: a preliminary study. In: Proceedings of the CECIIS 2023, 35th Central European Conference on Information and Intelligent Systems. Dubrovnik, Croatia, pp. 43–48.

Bohanec, M., 2025. DEXi Suite: DEXi decision modelling software. SoftwareX 31, 102240. https://doi.org/10.1016/j.softx.2025.102240.

Campanella, G., Ribeiro, R., 2011. A framework for dynamic multiple-criteria decision making. Decis. Support Syst. 52 (1), 52–60. https://doi.org/10.1016/j.dss.2011.05.003. December2011.

Cash, T.N., Oppenheimer, D.M., 2025. Assessing metacognitive knowledge in subjective decisions: the knowledge of weights paradigm. Think. Reason. 31 (3), 331–373. https://doi.org/10.1080/13546783.2024.2426543.

Cheng, X., Xu, Z., Gou, X., 2024. A large-scale group decision-making model considering risk attitudes and dynamically changing roles. Expert Syst. Appl. 245. https://doi.org/10.1016/j.eswa.2023.123017.

Deguine, J.P., Robin, M.H., Camilo Corrales, D., Vedy-Zecchini, M.A., Doizy, A., Chiroleu, F., Quesnel, G., Païtard, I., Bohanec, M., Aubertot, J.N., 2021. Qualitative modeling of fruit fly injuries on chayote in Réunion: development and transfer to users. Crop Prot. 139. https://doi.org/10.1016/j.cropro.2020.105367.

Edland, A., Svenson, O., 1993. Judgment and decision making under time pressure: studies and findings. In: Svenson, O., Maule, A.J. (Eds.), Time Pressure and Stress in Human Judgment and Decision Making. Plenum Press, pp. 27–40. https://doi.org/10.1007/978-1-4757-6846-6_2.

Fischhoff, B., Kadvany, J., 2011. Risk: A Very Short Introduction. Oxford University Press.

Fischhoff, B., Broomell, S.B., 2020. Judgment and decision making. Annu. Rev. Psychol. 71 (1), 331–355. https://www.annualreviews.org/doi/pdf/10.1146/annurev-psych-010419-050747.

Greco, S., Ehrgott, M., Figueira, J., 2016. Multi Criteria Decision Analysis: State of the art Surveys. Springer, New York. https://doi.org/10.1007/978-1-4939-3094-4.

Jackson, P., 1998. Introduction to Expert Systems, 3rd ed. Addison-Wesley.

Kahneman, D., 2003. A perspective on judgment and choice: mapping bounded rationality. Am. Psychol. 58 (9), 697–720. https://doi.org/10.1037/0003-066X.58.9.697.

J. Korhonen, P., Silvennoinen, J., Wallenius, K., Öörni, A., 2012. Can a linear value function explain choices? An experimental study Eur. J. Oper. Res. 219 (2). https://doi.org/10.1016/j.ejor.2011.12.040.

Kourouxous, T., Bauer, T., 2019. Violations of dominance in decision-making. Bus. Res. 12, 209–239. https://doi.org/10.1007/s40685-019-0093-7.

Kulkarni, A.J., 2022. Multiple criteria decision making. Studies in Systems, Decision and Control 407. Springer, Singapore. https://doi.org/10.1007/978-981-16-7414-3_3.

Leondes, T.C., 2002. Expert Systems: The Technology of Knowledge Management and Decision Making for the 21st Century, 1st ed. Academic Press.

Morris, A., Carlson, R.W., Kober, H., Crockett, M.J., 2025. Introspective access to value-based multi-attribute choice processes. Nat. Commun. 16, 3733. https://doi.org/10.1038/s41467-025-59080-y.

Moscati, I., 2023. The History and Methodology of Expected Utility. Cambridge University Press. https://doi.org/10.1017/9781009198295.

Nisbett, R.E., Wilson, T.D., 1977. Telling more than we can know: verbal reports on mental processes. Psychol. Rev. 84 (3), 231–259. https://doi.org/10.1037/0033-295X.84.3.231.

Novella, R., Ramirez, E.G.R., 2024. Question-order effects on judgements under uncertainty. J. Behav. Exp. Econ. 109. https://doi.org/10.1016/j.socec.2023.102159.

Phillips-Wren, G., Adya, M., 2020. Decision making under stress: the role of information overload, time pressure, complexity, and uncertainty. J. Decis. Syst. 29, 1–13. https://doi.org/10.1080/12460125.2020.1768680.

Raileanu, L.E., Stoffel, K., 2004. Theoretical comparison between the Gini index and information gain criteria. Ann. Math. Artif. Intell. 41 (1), 77–93. https://doi.org/10.1023/B:AMAI.0000018580.96245.c6.

Rezaei, J., Arab, A., Mehregan, M., 2021. Equalizing bias in eliciting attribute weights in multiattribute decision-making: experimental research. J. Behav. Decis. Mak. 35 (2). https://doi.org/10.1002/bdm.2262.

Rokach, L., Maimon, O., 2015. Data Mining with Decision Trees: Theory and Applications. World Scientific, New Jersey. https://doi.org/10.1142/9789812771728_0001.

Saaty, T.L., Vargas, L.G., 2012. Models, Methods, Concepts & Applications of the Analytic Hierarchy Process. Springer, New York. https://doi.org/10.1007/978-1-4614-3597-6.

Santos, C., Cunha, P., 2021. Influence of trust, time pressure and complexity factors in judgment and decision-making in auditing BBR. Braz. Bus. Rev. 18, 605–623. https://doi.org/10.15728/bbr.2021.18.6.1.

Silva, F.F., Souza, C.L.M., Silva, F.F., et al., 2021. Elicitation of criteria weights for multicriteria models: bibliometrics, typologies, characteristics and applications. Braz. J. Oper. Prod. Manag. 18 (4). https://doi.org/10.14488/BJOPM.2021.014.

Smoliński, P.R., Brycz, H., 2024. Individual differences in inaccurate versus accurate economic judgment and decision making. Metacognitive Approach 219. https://doi.org/10.1016/j.paid.2023.112500.

Taherdoost, H., Madanchian, M., 2023. Understanding applications and best practices of DEMATEL: a method for prioritizing key factors in multi-criteria decision-making. J. Manag. Sci. Eng. Res. 6 (2), 17–23. https://doi.org/10.30564/jmser.v6i2.5634.

Thakkar, J.J., 2021. Multi-criteria decision making. In: Studies in Systems, Decision and Control, 336. Springer, Singapore. https://doi.org/10.1007/978-981-33-4745-8.

Tversky, A., Kahneman, D., 1992. Advances in prospect theory: cumulative representation of uncertainty. J. Risk. Uncertain. 5, 297–323.

Vetschera, R., Weitzl, W., Wolfsteiner, E., 2014. Implausible alternatives in eliciting multi-attribute value functions. Eur. J. Oper. Res. 234 (1), 221–230. https://doi.org/10.1016/j.ejor.2013.09.016.

Vohs, K.D., Luce, M.F., 2010. Judgment and decision making. Advanced Social Psychology: The State of the Science. Oxford University Press.

Yang, S., Liao, L., Xingli, W., 2025. Prescriptive analytics for dynamic multi-criterion decision making considering learned knowledge of alternatives. Expert Syst. Appl. 268. https://doi.org/10.1016/j.eswa.2024.126350.