Decision Noise Instrument (DNI): Estimating Decision Noise in Business Processes

Marcus Dees[1], Peter H.G. Berkhout[2], Claudio Di Ciccio[3], and Hajo A. Reijers[3]

[1] Radboudumc, Nijmegen, The Netherlands, marcus.dees@radboudumc.nl
[2] UWV, Amsterdam, The Netherlands, peter.berkhout@uwv.nl
[3] Utrecht University, Utrecht, The Netherlands, {c.diciccio,h.a.reijers}@uu.nl

Abstract. Business processes in which decisions are made by human resources suffer from noise. Noise is unwanted variability, which leads to inconsistent and unrepeatable decisions and impacts trust negatively. In this paper, we present the Decision Noise Instrument (DNI) to quantitatively estimate the noise that can be attributed to unknown factors affecting the actors who carry out the business process. For our estimation, we solely use readily available data from business information systems that support the execution of the process. In our method, we limit the influence of factors that cannot be attributed to actors. The DNI makes it possible to compare the noise levels between and within business processes. Based on that comparison, further investigation into the problem and noise reduction efforts can be prioritized. We evaluate the DNI on a claim handling process of UWV, a public service provider in the Netherlands. Our results show that we can estimate differences between the noise levels of several decisions within the selected process of UWV. By drilling down on a decision, the subgroups that have the most impact on the noise level are identified.

Keywords: Business Process Management · Decision analysis · Noise

1 Introduction

Decision making is an integral part of all business processes [19,21]. Examples of processes in which decisions play a key role are diagnostic processes operated by doctors, legal processes executed by judges, and insurance claim processes carried out by insurance experts. These types of processes also have in common that the final decision is typically left to a human decision maker, since they are trusted and expected to make a fair, objective decision. We refer to these types of processes as *decision processes*. Especially for decisions that have a major impact, it is essential that the decision is trusted by all parties who are involved in and subject to the decision process. Variability is in itself not inherently good or bad. The variation that occurs within a decision process should, however, be explainable and made accountable. If this cannot be guaranteed then trust in the process itself and the organization executing the process deteriorates.

Noise and bias are the core causes for variability within decision processes, whereby bias is a systematic variation is in a specific direction, and noise is unexpected variation. Kahneman et al. [13] show that human decision makers are unaware of *noise* in their decisions, which means that under the *same observable circumstances* a human actor does not always reach the *same decision*. A part of the variability of decision process can therefore be attributed to decision makers. We use the term *decision noise* when we refer to noise that originates from taking a decision by a human actor, i.e., when the human actor is the source of the noise. In the literature, this is also called *system noise* [13]. The effect of having decision noise in a business process means that not only the behavioral properties or the norms regulating the decision process have an impact on the outcome of the decision, but also *who* takes the decision ultimately matters. The state-of-the-art intervention proposed in [13] is to estimate the overall amount of noise in a decision process by performing a *noise audit*. In a noise audit, a group of resources assesses the same pre-compiled cases. Comparing their decisions allows for estimating the variability in the decision process. The outcome of a noise audit can uncover the amount and the sources of decision noise. However, performing it is no sinecure, as Kahneman et al. suggest in [13, Appendix A].

In light of the above, we present the Decision Noise Instrument (DNI) as a means to estimate decision noise in business processes based on process data. Our method is less labor-intensive than a noise audit, and uses data about decision-making that is already recorded in the information systems supporting the execution of business processes. Our objective is to *automate* the estimation of noise levels between or within decision processes. We use a statistical behavioral model [5] in combination with a simulation model to estimate the decision noise level of a decision process or the *decision checks* that make up a decision process. Our method approximates the *true decision noise* that can be attributed to the human actors by limiting the influence of factors that cannot be attributed to human actors. To gauge the noise level of a decision process, we introduce a measure we call *switching rate*. The estimated noise levels can be compared in order to prioritize further research and efforts to reduce noise within business processes. We perform a case study applying the DNI in the context of the unemployment benefits claim process at the Dutch Employee Insurance Agency UWV (Uitvoeringsinstituut Werknemersverzekeringen). The case study shows that our method is able to estimate the noise levels of different decision checks made within the selected decision process. The results reportedly gave UWV the opportunity to target further investigation in the most noisy decision checks.

The remainder of the paper is structured as follows. In Sect. 2, we discuss related work. In Sect. 3, we explain the method we developed. Section 4 details the design of our statistical technique and how a dataset can be constructed as an input. In Sect. 5, we report on a case study at UWV using our method. Section 6 discusses the generalizability and limitations of our work. Section 7 concludes the paper with directions for future work.

2 Related Work

Our work is directly inspired by the book *Noise: A Flaw in Human Judgement* by Kahneman et al. [13]. Variation exists in human judgment and in business processes executions by human resources. Variability within a process is in itself not a negative property. It can even be a design choice of the process and of the organization [18]. Noise, however, is a source of variability that organizations, in general, should strive to minimize. This is especially true for decision processes. It is the aim of a decision process to create accurate and repeatable decisions *independently* of the human actor taking the decision. That is, from the perspective of the subject in the decision process the outcome of the decision should not depend on the assigned actor.

The influence and impact of decision noise have already been investigated in other domains such as healthcare and sports. In the healthcare domain, noise affecting the decision on the diagnostic test to be performed is investigated by Belle et al. [3] and Dlugos et al. [7]. In the sports domain, noise in the decision-making of football referees is examined by Samuel et al. [17].

Our work aligns with the notion of *fairness* in Artificial Intelligence [2,9], which also seeks to distinguish justified from unjustified variability in decision making. Decision noise analysis is, from this viewpoint, similar: We aim at a model that helps to separate variability into *systematic variability*, which is predictable and explainable, and *idiosyncratic variability*, i.e., noise.

Understanding how noise can be modeled in computational cognitive models is discussed by Li et al. in [14]. The authors state that most models use a static noise level. Better models for decision-making can be obtained by modeling noise in a dynamic way. Li et al. argue that periods of attention to the decision task are alternated with moments of distraction. While our aim is to measure the actual noise level of a decision process, understanding how to model noise can prove to be helpful in designing noise reduction methods.

Recommender Systems (RSs) [1] also have to be able to deal with noise. To train a RS, samples from rating databases are used that can contain *natural noise* and *malicious noise* [15]. Natural noise originates from users not being consistent in giving ratings. A source of malicious noise lies in ratings that are deliberately added to steer a rating in a specific direction, thus creating bias. For RSs, countermeasures have been proposed to deal with noise, including noise-reducing sampling and noise-filtering methods [11]. Some of the techniques discussed could also be viable ways to reduce noise in a business process context.

Bollarino et al. [4] developed a *noise decision scale* aimed at examining organizational factors that have an influence on decision-making and the level of noise. The authors use qualitative surveys to develop and validate their scale. Similarly, in [20], research on reducing noise is presented and different strategies are tested. A survey was used to gather data about the effectiveness of each noise reduction strategy. In our method, we refrain from resorting to labor-intensive approaches like surveys to estimate decision noise levels, and opt for a data-based measurement, as we explain in the next section.

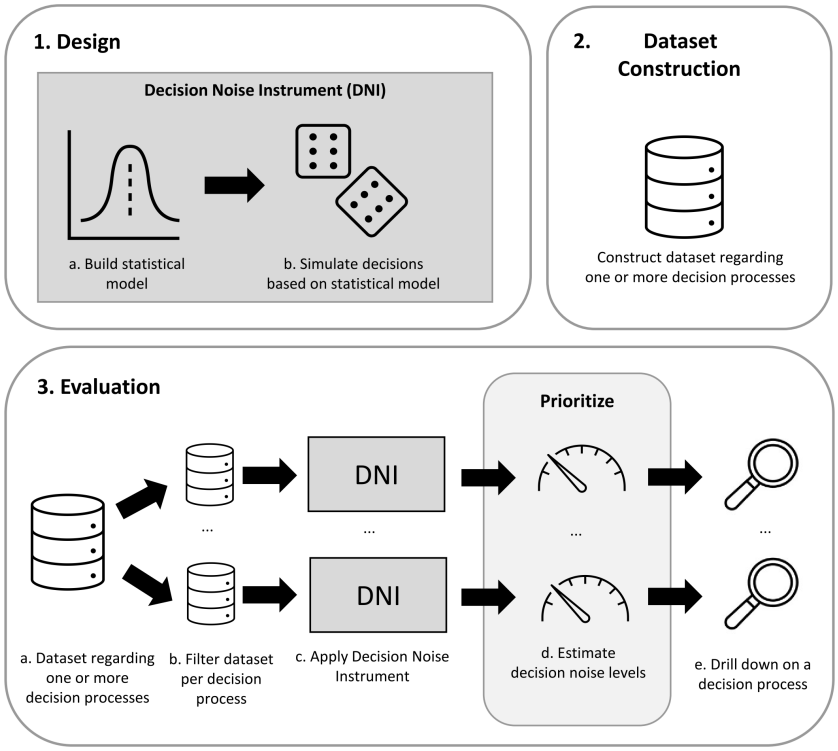

Fig. 1: Overview of the three main phases of our method

3 Method

The most important countermeasure to manage noise, as suggested by Kahneman et al. [13], is to perform a noise audit. The goal of a noise audit is to calibrate a decision process by having several human resources judge the same set of pre-compiled cases. By comparing the judgments of the human resources for each case, the amount of noise in the decision process can be estimated. The noise audit, as proposed in [13], is a manually-intensive exercise. In our work, we propose an automated approach instead. It applies to cases wherein data from the decision process is available, and the assignment of actors to cases is random.

Figure 1 gives an overview of the method we employed to conduct our research. In the first step, we designed the Decision Noise Instrument (DNI). The DNI is an instrument that estimates decision noise based on a dataset taken from the decision process. The DNI combines (a) a multilevel mixed-effects model [5] with (b) a simulation model [10]. Intuitively, the DNI returns the percentage of cases whose decision is expected to change when handled by a different human actor. Secondly, we describe the requirements to construct a dataset that can be used together with the DNI. In the third and final step, we evaluate our method by performing a case study. In particular, we employed the DNI in the

real-world setting of the unemployment benefits claim decision process at UWV. We started the evaluation with (a) the extraction and cleansing of the dataset. Next, (b) we split the dataset into separate chunks, one for each *decision check*. A decision check represents an individual decision within the decision process with its own nature and complexity, and a dedicated decision handling process. For example, a claim on the unemployment benefits process can require a check on the *Weeks requirement*, which states that a claimant needs to have had an income in 26 weeks out of the 36 weeks prior to unemployment. For another claim, the *Notice period* may need to be checked: an employee cannot become unemployed between the date that the employer gives the notification of dismissal and the contractually agreed period that the employer must adhere to. Although the above decision checks seem rather automatic, empirical evidence shows that the outcome varies depending on the resource handling it. The variation can, to some extend, be explained by the discretionary space the UWV resource has. This means that for each case and check properties and context of the claimant's request need to be weighed in.

A decision check is also a separate (sub-)decision process on its own. Therefore, our method also allows us to make a comparison *between* multiple decision processes. The DNI (c) is applied on each dataset to obtain estimations of the decision noise level for each decision check. We compare the estimated decision noise levels per decision check (d) to guide further investigation and noise reduction efforts. For each decision check, different subgroups of claims may exist with different decision noise levels. We drill down (e) on each decision check to determine the most crucial subgroups and better steer the forthcoming efforts.

4 Design

The aim of the DNI is to quantify and thereby compare decision noise levels between decision processes, or decision checks within a decision process. In Sect. 4.1, we discuss the design of the DNI. In Sect. 4.2, we explain our approach for the construction of a suitable dataset on which the DNI can be applied.

4.1 The Decision Noise Instrument

The DNI is developed based on the idea that a decision process with a binary outcome can be expressed as a statistical behavioral model. A behavioral model captures the essence of the decision-making process of human resources in a statistical model. This is the first element (1a) of the DNI as displayed in Fig. 1.

Statistical Behavioral Model. We aim to estimate the noise level of a decision process executed by human resources. We assume that each case i for which a decision needs to be made has an objective (and unknown) grant-ability value $G_i \in \mathbb{R}$. Human resources working in the process make their own estimation of the value of G_i. The decision for the case is either 1 (denoting a positive outcome, like *Yes*, *Accepted*, *OK*, etc.), when the estimation by the human resource corresponds

to $G_i \geqslant 0$, or 0 (indicating a negative outcome: *No*, *Rejected*, *NOK*, etc.), when $G_i < 0$. If every human resource could estimate the same value of G_i for each case i, the decision process would have no decision noise. However, when two resources make their own estimation and their estimates differ from each other, then the same case i could be associated to $G_i = 1$ for the first resource and $G_i = 0$ for the second resource. The source of variation in this model is the extent to which an individual's estimate deviates from the actual value G. The real value for G is unknown, though, since it is a concept that solely exists within our model. We now assume that G_i of case i is equal to the hypothetical average if all human resources were to make an estimate of it. Let R be the number of resources. The index $r \in [1, R]$ identifies a unique resource. Each resource r has a (structural) deviation T_r. We assume that this deviation is the same for every case processed by r. We postulate that T_r follows a normal distribution with the average equal to 0 and the standard deviation set to σ_T: $T_r \sim \text{Normal}(0, \sigma_T)$. A negative T_r means that r structurally grants cases less often than the average. We assume that the objective grant-ability G_i of case i follows the logistic distribution: $G_i \sim \text{Logistic}(X_i\beta, 1)$. The first argument (the location parameter), can vary depending on application-specific features X_i which are typically collected in a vector and multiplied by a vector of coefficients β, such that the features can each have their own weights. A feature represents a source of variation. An example of a feature is the age of the applicant if, in practice, older people are awarded more often (or less often) than younger people on average. Vector X_i also always includes a constant, which represents the intercept in the regression equation. We set the scale argument to 1 since this assumption does not change the essence of the model.

The decision $D_{i,r}$ of resource r on case i is thus determined as follows:

$$D_{i,r} = \begin{cases} 0, & \text{if } G_i + T_r < 0 \\ 1, & \text{if } G_i + T_r \geqslant 0 \end{cases} \tag{1}$$

To build our statistical model, we need records of the following data: a case i with its properties X_i, a resource r, and the actual decision $D_{i,r} \in \{0, 1\}$ taken by r for case i. We do not observe the variables G and T. The objective is to estimate the model parameters β and σ_T from the available data. Multilevel mixed-effect models are suitable for analyzing this sort of information. More specifically, the described model can be estimated empirically with a logit model with fixed effects for case properties and a random effect for the human resources.

Simulation. The resulting model estimate of σ_T is an estimate of the standard deviation of the individual deviations of the human resources. The value of σ_T expresses the variability in decision making for all human resources that execute the process. That is, σ_T is a benchmark for decision noise within a process that is related to the human resources executing the process. This implies that σ_T could already be used to benchmark decision noise levels.

Domain experts involved in our study expressed a preference towards a more tangible measure than standard deviations. To this end, we introduced

the *switching rate* S as a measure of decision noise. We define the switching rate as the percentage of cases that are expected to change outcome in the event that the case were assigned to another randomly assigned resource. The percentage of cases whose outcome depends on the handling resource can be seen as the degree of arbitrariness of the system. A low percentage indicates little arbitrariness, whereas a high percentage indicates a high level of arbitrariness. The *switching rate* $S(\sigma_T, \bar{X}\beta)$ follows from the estimated values for σ_T and $\bar{X}\beta$ and the assumed distributions of the model (i.e. the logistic distribution and the normal distribution). We define S as follows:

$$S(\sigma_T, \bar{X}\beta) = \iint_{-\infty}^{+\infty} |\Lambda(r_1) - \Lambda(r_2)| \cdot \Phi(r_1) \cdot \Phi(r_2)\, dr_1\, dr_2. \tag{2}$$

In the formula, Λ is the cumulative distribution function of the logistic distribution with expectation $\bar{X}\beta$ and scale 1, and Φ is the probability density function of the normal distribution with expectation 0 and standard deviation σ_T. The integral calculates the difference between two human resources making a decision for the same case. Inspired by Item Response Theory [16], we use it as a means that is also common in psychometric models to measure noise or discrimination.

We approximate $S(\sigma_T, \bar{X}\beta)$ by running a simulation using standard computation methods for the normal distribution and the logistic distribution [10]. Figure 2 shows a contour plot indicating the value $S(\sigma_T, \bar{X}\beta)$ for each combination of the standard deviation σ_T and expected acceptance rate $\Lambda(\bar{X}\beta)$. The maximum switching rate is 50 %. When the expected acceptance rate is close to 0 % (or 100 %) the switching rate is close to 0 %. The statement that the switching rate is bounded by 50 % may seem counterintuitive. One might imagine two decision makers who always make opposite choices, implying 100 % disagreement. However, the switching rate as defined here refers to the probability of disagreement when two decision makers are drawn at random from the population. Technically, let p be the probability of a positive decision. In this setting, the disagreement probability is $2p(1-p)$. This expression is maximized at 0.5, yielding an upper bound of 50 %.

4.2 Dataset Construction

The DNI uses a dataset taken from a decision process as input. Both the decision process and the dataset need to meet specific requirements.

Decision process requirements. To avoid variation between human resources that originates from the process itself, we require that the human resources are *randomly assigned* to the decisions that need to be taken. This ascertains that no assignment bias occurs in the dataset. When some bias purposely exists within a process, for example because certain human resources are always assigned a specific type of cases, then the DNI should be applied separately on each set of human resources that deal with the same type of cases.

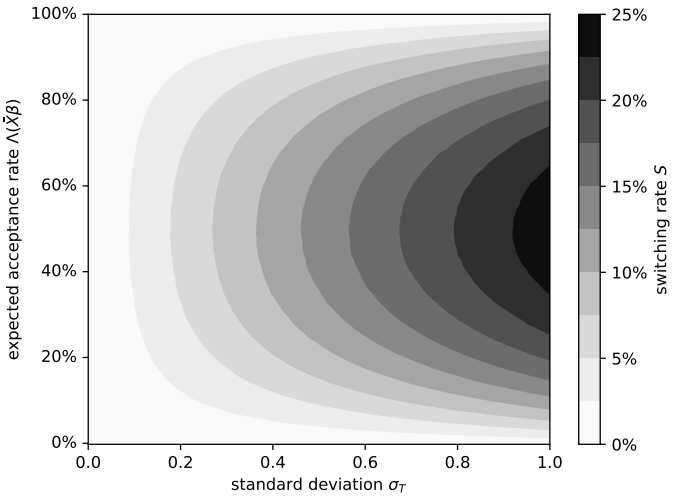

Fig. 2: Contour plot showing the percentage of cases that are expected to change at specific acceptance rates and standard deviation values.

The mix of cases each human resource is assigned is another source of variation. While the cases have the same type, differences between the cases can still exist. To minimize this type of variation, we advocate, as a starting point, that per human resource at least 20 cases per decision process should be present. In our evaluation, we observed that having at most 15 resources in the dataset is not enough to derive a valid statistical model. We advise having at least 30 different resources in the dataset. This implies that the dataset should consist of at least 600 cases. Our approach is, therefore, only feasible for high-volume decision processes, i.e., when enough human resources have handled a sufficiently large number of cases.

Another requirement for the decision process is that the decisions' outcome should have exactly one of two possible values. For example, a decision process that handles an insurance claim has *accepted* and *rejected* as possible outcomes. In our approach, we calculate the share of cases that is expected to switch outcome, i.e., from accepted to rejected or vice versa.

Dataset requirements. The dataset should contain at least the following pieces of information: an attribute that identifies a case, an attribute that identifies a human resource, and a decision, having the value 0 or 1 to indicate the binary outcome. In general, the estimation of the decision noise will become more accurate if more sources of variation are excluded from the dataset.

For each decision process, specific sources of decision noise should be identified and removed from the dataset if possible. In practice, this can turn out to be difficult. Also, not all sources of variation can be captured in data. Our approach is, however, robust enough to deal with missing sources of variation. The estimation

in itself may become less accurate as an approximation of the true noise. However, it is still possible to compare estimates with each other. Our technique is able to identify the decision process with the biggest amount of noise, as long as the same sources of variation are present within each dataset per decision process and are used in the statistical model.

To be able to drill down on decisions, more context-specific variables should be added to the dataset. After the comparison of the decision noise levels between decision processes, these variables can be used to distinguish differences between subgroups of cases. Understanding which subgroups exist will support in focusing further research and noise reducing efforts. In the next section, we will evaluate both the construction of a dataset and the DNI for a real-life decision process.

5 Case study

UWV (Uitvoeringsinstituut Werknemersverzekeringen)[4] is a public service provider in the Netherlands. In particular, it is responsible for supporting citizens to get back to work. If this is not immediately or fully possible, UWV provides a temporary income in the form of social security benefits.

Recently, UWV introduced a new guideline to cater for more customization of the decision processes. In particular, the UWV resources would be given more discretionary space to tailor decisions to client-specific situations in a more fitting way. As per the new guideline, the UWV resource should weigh in the situation of the citizen and not solely rely on the rule of law. Thus, the new guideline increases the influence of the UWV resource on each individual decision and thereby decreases the level of determinism of a decision. As an expected consequence, the presence of decision noise might increase.

However, since UWV did not have an instrument to gauge the decision noise that is currently present in their processes, it was not possible for them to determine whether the new guideline would actually raise the noise levels or not. In an attempt to tackle this problem, UWV agreed to run a pilot study with the application of the DNI on one of their claim processes to estimate the magnitude of the decision noise present in it.

5.1 The unemployment benefits claim process of UWV

Employees in the Netherlands can claim unemployment benefits at UWV. Figure 3 shows a simplified BPMN model of the process for the handling of a claim up until reaching a decision. The client who files a claim is referred to as the *claimant*. The claim is received in the workflow system of UWV (*Register claim*). Information about former employments and income of the claimant is automatically gathered from other governmental agencies like the taxing service and added to the claim. If all required information is present to make a decision on the claim, then the decision is made automatically (*Take automatic decision*). Otherwise, the

[4] https://www.uwv.nl/. Accessed: June 12, 2025.

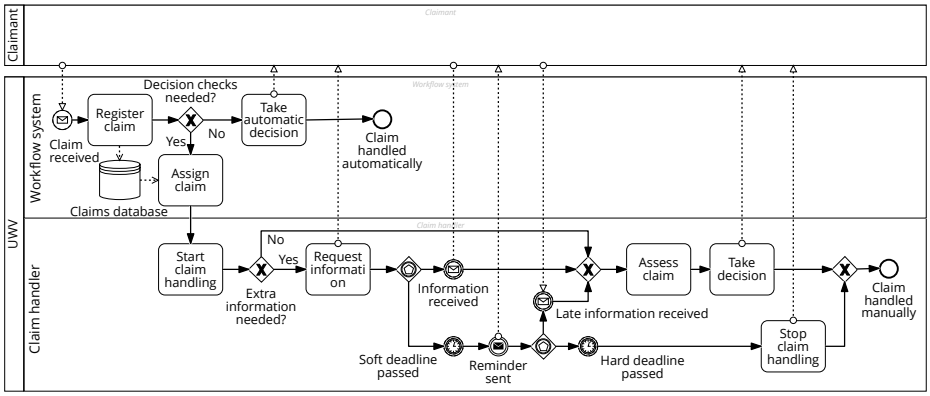

Fig. 3: BPMN model of the unemployment benefits claim process at UWV

workflow system assigns the claim to a randomly selected resource at a UWV office (*Assign claim*). When necessary, extra information is requested that is not part of the claim (e.g., the employment contract of the claimant). When the information is not received within a week, a reminder is sent. If no information is received after two weeks, then the claim handling is stopped (*Stop claim handling*). As soon as the necessary pieces of information are acquired, the UWV resource assesses the claim (*Assess claim*) and takes a decision (*Take decision*) which is delivered to the claimant.

To reach a final decision, one or more decision checks need to be made by a UWV resource. Which decision checks need to be done is automatically derived from the information that is present in the received claim. Examples of decision checks are the *Weeks requirement* and the *Notice period*, which we mentioned in Sect. 3. Other decision checks, with a brief description, are:

- *Interim termination*: the contract is stopped before the contractual end date;
- *Daily wages*: information to calculate benefits amount is missing;
- *No relevant loss*: a claimant must have lost at least 4 working hours per week;
- *Legitimate claim*: did the claimant or the employer cause the unemployment?

The UWV resource who is assigned the handling of the claim is informed about which decision checks need to be done. Each decision check has its own procedure and can be considered a distinct decision process. If one of the decision checks leads to a rejection, the claim itself is rejected. Otherwise, the claim is accepted. As mentioned before, within each case and check, the UWV resource has the discretionary space to weigh in properties and context of the claimant's request. For each of the decision checks, then, we want to estimate the noise level.

5.2 Application of the DNI at UWV

To apply the DNI on the unemployment benefits claim process at UWV, we constructed a dataset based on claims that were received between August 2021 and September 2024. The first step was to remove as many potential sources

Table 1: Sample of four cases and some of the variables from the dataset. The decision checks indicate which check needs to be done for the claim. The decision and other boolean variables are encoded as 1 for *true/yes* and 0 for *false/no*.

Variable	**Claim 1**	**Claim 2**	**Claim 3**	**Claim 4**
Decision: Accepted	1	1	0	1
Resource	R1	R2	R3	R1
Office	A	B	A	E
Age of claimant	37	55	50	62
Contract type: Permanent	1	0	1	0
...				
Decision check: Weeks requirement	1	1	0	1
Decision check: Notice period	0	0	1	0
...				

of variation from the dataset. We did that by filtering out claims for which no decision could be made, i.e., only claims that were accepted or rejected were selected. Next, we retained claims for which exactly one UWV resource was involved in the handling of the claim. Claims that were handed over between colleagues or between offices are not used because multiple resources were involved in taking the decision. Finally, a claimant can file multiple claims during the research period. Only the first claim for each claimant was used. This ensures that any previous knowledge about a claimant is minimized.

Table 1 shows an excerpt of the dataset. The outcome of the decision is assigned with either of two values, which we encoded as 1 for *accepted* and 0 for *rejected*. After the removal of claims that do not meet the described criteria, we obtained a dataset containing 549,772 claims, handled by 1739 unique UWV resources. The average acceptance rate in the research dataset is 85.7 %. Differences exist between the offices where the claims were handled, though, with acceptance rates ranging from 83.0 % to 87.8 %.[5]

As mentioned in Sect. 4.1, independent variables in the regression model are split into fixed factors that are not considered sources of decision noise and random factors which are seen as sources of decision noise. Factors that represent decision noise are factors that are related to how the process is designed and therefore are an integral part of the process. For the selected claim process, these factors are the UWV resource that executes the process and the office where the UWV resource resides. In our dataset, the moment when a claim was received is also a source of decision noise since the research period that we consider spans a period of 3 years. In this time frame, the process itself has changed and the UWV resources who work in the process have changed, too.

We applied the DNI using only the factors that represent decision noise, i.e., the UWV resource, the UWV office and the year and month when the claim was received. These factors were added as random effects to the regression expression. No other variables (i.e., fixed effects) were used. For each decision check a separate

[5]The acceptance rates are based on claims with only one decision check.

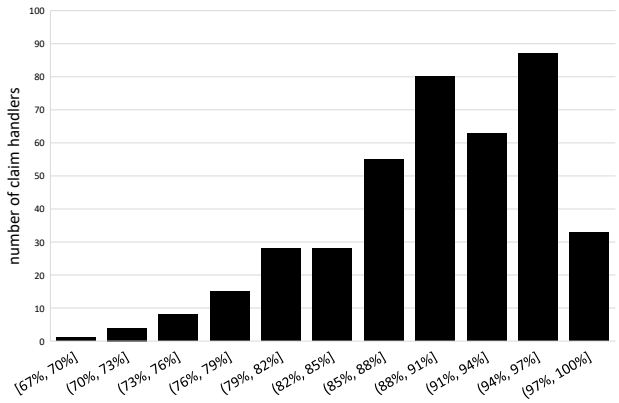

Fig. 4: Distribution of the acceptance rate of UWV resources who handled at least 20 claims subject to the *Weeks requirement* decision check

Table 2: Summary of the UWV dataset

Decision check	# claims	# resources	Acceptance rate	Standard deviation	Switching rate
Weeks requirement	12,892	407	89.8%	0.337	3.4%
Notice period	3393	123	74.6%	0.535	11.0%
Interim termination	2625	102	85.4%	0.339	4.7%
Daily wages	18,297	525	95.3%	0.429	1.7%
No relevant loss	12,071	355	97.1%	0.555	1.7%
Legitimate claim	4505	169	78.4%	0.308	5.8%

dataset was constructed by filtering the complete dataset. We built our statistical model using *Pymer4* [12], a Python implementation of the R package *Lme4*.

Taking the *Weeks requirement* decision check as an example, we filtered out from the dataset the claims that were handled by resources who were not involved in at least 20 claims requiring that check. This ensures that we have enough samples for each resource in the dataset for this decision check. After filtering, the dataset contains 12,892 claims. The average acceptance rate within the filtered dataset is 89.8 %. The dataset holds 407 unique resources from different UWV offices. Figure 4 shows the distribution of the acceptance rate of resources for claims where only the *Weeks requirement* decision check was conducted. The range for the acceptance rate for the 407 UWV resources is between 67 % and 100 %. The oscillation is evidently wide. We investigated to what extent this phenomenon could be attributed to decision noise. Next, we report our findings.

5.3 Results

We applied the DNI on all decision checks. Table 2 summarizes the results we gathered. For each decision check, it shows the number of claims used in the analysis, the number of unique resources involved, the acceptance rate of these claims, the standard deviation σ_T, and the estimated switching rate S. Notice

Table 3: Subgroups for the *Weeks requirement* based on the variable *Contract type* and the estimated decision noise levels

Subgroup	# claims	# resources	Acceptance rate	Standard deviation	Switching rate
Contract type: Temporary	3254	123	89.7%	0.397	4.1%
Contract type: Permanent	581	22	99.0%	0.919	2.1%

that a comparison among decision noise levels is legit because we estimated them under the same experimental conditions. As we can see in Tab. 2, the *Notice period* decision check exhibits the highest decision noise level at 11.0 %. The acceptance rate of this decision check is 74.6 % and the standard deviation of the random effect for the UWV resource is 0.535. As can be observed in Fig. 2, these values correspond with a switching rate between 10.0 % and 12.5 %. Next, we drill down on one decision check to uncover relevant subgroups.

5.4 Drilling down on a decision check

Any variable present in the dataset can be used to drill down on, as long as the variable is not part of the integral design of the system and thereby may signify decision noise. Since subgroups are per definition smaller than the complete dataset, it is possible we do not have enough cases for each subgroup to derive the noise level. As mentioned in Sect. 4.2, we need roughly 600 cases to estimate the decision noise level. We use the dataset for the *Weeks requirement* decision check again. The variable *Contract type* is part of the dataset. This variable represents the type of contract that the claimant has with an employer. A claimant has either a permanent contract, or a temporary contract. For each value, i.e., subgroup, the average acceptance rate and the noise level is estimated and shown in Tab. 3.

The subgroup *Contract type: Temporary* has almost a twice as high decision noise level (4.1 %) compared to the subgroup *Contract type: Permanent* (2.1 %). Keeping Fig. 2 in mind, the expected acceptance rate plays an important role. At 99.0 % the subgroup *Contract type: Permanent* has a 10 % higher acceptance rate than the subgroup *Contract type: Temporary* (89.7 %). The number of claims and resources involved, together with the decision noise level of 4.1 % make it clear that it is best to focus first on further qualitative and quantitative research investigations on the subgroup *Contract type: Temporary*.

The next section will address UWV's response to our analysis, alongside implications for practice and research, and present limitations of our method.

6 Discussion

Having DNI at our disposal, organizations can now benefit from an approach, other than the noise audit, to measure decision noise in their decision processes. However, applying the DNI does come at a cost and carries a responsibility.

Implications for practice. When the results of the DNI analysis were presented to UWV, the first question raised during the follow-up meeting was which noise level is acceptable from a policy point of view for each decision check. Determining a threshold of this sort is up to the stakeholders, and further case studies are necessary to start building up guidelines in that regard. Organizations should be aware that estimating decision noise will trigger questions like these, and thus should be required to formulate maximum noise level policies.

Another consequence of our approach is when the DNI estimates the switching rate at for example 6 %, then 6 out of 100 cases are expected to have a differing outcome whenever another human resource handles them. This raises the question of whether an organization should actively search for cases that could have had a different outcome. Apart from the practical hurdles of setting up such an investigation, an ethical question arises, too: What should an organization do when these cases are found? Stakeholders should decide what to do before starting the endeavor. In fact, understanding that this ethical aspect of our method exists ahead of time is crucial. Otherwise, it might hinder further consideration.

Implications for research. Decision noise is present in business processes where human actors take decisions. Other types of processes are suitable for research on decision noise. An illustrative example is the peer-review process in academic publishing. This process is compelling due to the fact that manuscripts are typically evaluated by three or more independent reviewers, allowing for robust analysis of inter-rater variability.

The decision processes we investigate are rooted in laws, policies, and work instructions. The unemployment benefits process from our case study is supported by a workflow system and a rule-based system, which is similar to a decision model expressed via Decision Modeling and Notation (DMN)[6] [8,6]. Although DMN can be used to mitigate decision noise through standardization, transparency, automation, and validation of the decision process, decision noise affects the decision process as soon as the latter leaves room for arbitrary interpretation.

Other techniques may complement and enrich our approach. Machine Learning, e.g., could be used to estimate decision noise by modeling the systematic component of decisions, analyzing residuals or unexplained variance, comparing human versus model consistency, and using statistical decomposition (variance components, entropy) to quantify randomness. Ideally, alternative decisions could be predicted for every observation in a sample by swapping human resources, thus obtaining a switch rate. Nevertheless, such achievements are attainable only in the presence of rich and high-quality training datasets.

Being able to estimate decision noise more easily also opens up the potential to create decision noise reduction methods. One such effort is supporting the decision-making process of a human resource with a decision support system. The system can make a suggestion, but the human resources must justify when they want to deviate from it. The registered justifications can then be used to improve the decision support system.

[6] https://www.omg.org/spec/DMN. Accessed: September 24, 2025.

Limitations. The DNI, evaluated at UWV, can be generalized to other domains and processes, provided the key conditions discussed in Sect. 4 are met.

However, the DNI has its own limitations. Unlike controlled experimental methods, where multiple assessors evaluate identical cases, DNI relies solely on observational data. This requires the assumption that distinct cases can be treated as sufficiently similar to warrant comparison. The validity of this assumption hinges on the quality and completeness of the available features: if important dimensions of case complexity are missing, the method risks conflating legitimate differences with noise. However, we remark that we assume cases to be randomly assigned to resources. This aspect permits us to hint at the law of large numbers. All unobserved positive and negative factors influencing the outcome should zero out, thereby not interfering with the higher level estimates.

It is worth mentioning that the number of years of experience (i.e., tenure), and the skill set of resources might influence their decisions. If data of sufficient quality is available about these traits, then these factors could be added to the model as a source of variation.

At present, DNI cannot deal with cases that require multiple decisions (or decision checks). The DNI also does not indicate which cases have a higher probability of resulting in a varying outcome. Similarly, the DNI does not indicate which actors exhibit a higher level of decision noise than others. We built the DNI for decision processes with a discrete outcome and one resource involved. For other configurations, different models are needed, although multi-level mixed-effect models might still be suitable for many of those.

7 Conclusion

Noise is an important aspect of human decision making. Organizations should be aware of decision noise in their processes. Estimating human-induced noise in decision processes is a first step for organizations to enhance fairness and trust in their business processes. Our developed instrument (the DNI) has provided a national governmental agency such as UWV insights into the decision noise levels that exist within a claim process. We plan to gauge the effectiveness of the DNI in other organizations from different domains.

At present, DNI can only deal with binary decisions. However, other types of outcome values, i.e., nominal and continuous decision values, can still be modeled using the multilevel mixed-effects model we applied. For non-binary types of decisions, the instrument should be extended to estimate, instead of a switching rate, the percentage of the value that could change due to the noise level in the decision process. Since we aimed to automate the measurement of noise levels, our method can also be used to monitor the effect of any interventions that are deployed and follow the development of noise levels over time. In future work, we aim to develop a monitoring framework for decision noise levels.

Disclosure of interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Adomavicius, G., Tuzhilin, A.: Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE Trans. Knowl. Data Eng. **17**(6), 734–749 (2005)
2. Barocas, S., Hardt, M., Narayanan, A.: Fairness and Machine Learning: Limitations and Opportunities. Adaptive Computation and Machine Learning series, MIT Press (2023)
3. Belle, N., Cantarelli, P., Wang, S.Y.: The management of bias and noise in public sector decision-making: experimental evidence from healthcare. Public Management Review **26**(11), 3246–3269 (2024)
4. Bollarino, S., Ceschi, A., Monti, M., Sartori, R.: Development and validation of a noise in decision inventory for organizational settings. Informing Science: The International Journal of an Emerging Transdiscipline **26**, 173–189 (01 2023)
5. Bryk, A.S., Raudenbush, S.W.: Hierarchical linear models: applications and data analysis methods. Sage Publications, Inc (1992)
6. De Leoni, M., Felli, P., Montali, M.: Integrating bpmn and dmn: Modeling and analysis. Journal on Data Semantics **10**(1), 165–188 (2021)
7. Dlugos, K.V., Mazwi, M., Lao, R., Honjo, O.: Noise is an underrecognized problem in medical decision making and is known by other names: a scoping review. BMC Medical Informatics and Decision Making **25**(1), 86 (2025)
8. Dumas, M., La Rosa, M., Mendling, J., Reijers, H.A., et al.: Fundamentals of business process management, vol. 1. Springer (2013)
9. Dwork, C., Hardt, M., Pitassi, T., Reingold, O., Zemel, R.S.: Fairness through awareness. In: ITCS. pp. 214–226 (2012)
10. Forbes, C., Evans, M., Hastings, N., Peacock, B.: Statistical Distributions. Wiley (2011)
11. Jain, K., Jindal, R.: Sampling and noise filtering methods for recommender systems: A literature review. Engineering Applications of Artificial Intelligence **122** (2023)
12. Jolly, E.: Pymer4: Connecting R and Python for linear mixed modeling. Journal of Open Source Software **3**(31), 862 (2018)
13. Kahneman, D., Oliver, S., Sunstein, C.R.: Noise. A Flaw In Human Judgement. Springer (2021)
14. Li, J.J., Shi, C., Li, L., Collins, A.G.: Dynamic noise estimation: A generalized method for modeling noise fluctuations in decision-making. Journal of Mathematical Psychology **119**, 102842 (2024)
15. O'Mahony, M.P., Hurley, N.J., Silvestre, G.C.: Detecting noise in recommender system databases. In: IUI. p. 109–115 (2006)
16. Robitzsch, A., Steinfeld, J.: Item response models for human ratings: Overview, estimation methods, and implementation in r. Psychological Test and Assessment Modeling **60**(2), 223–243 (2018)
17. Samuel, R.D., Galily, Y., Hochman, G.: The "silent" noise: moving forward from bias to noise in football referees' decision-making. International Journal of Sport and Exercise Psychology pp. 1–26 (2025)
18. Shishkov, B., Mendling, J.: Business process variability and public values. In: Business Modeling and Software Design. pp. 401–411 (2018)
19. Simon, H.A.: Rational decision making in business organizations. The American Economic Review **69**(4), 493–513 (1979)
20. Srinivasan, S.: Quantifying Effects of Automated Noise Auditing Notices and Decision Structuring on Noise, Accuracy, and Fairness in Human Decision-Making. Master's thesis, Harvard University (2022)

21. Vanthienen, J.: Decisions, advice and explanation: an overview and research agenda. A Research Agenda for Knowledge Management and Analytics pp. 149–169 (2021)

This document is a pre-print copy of the manuscript
(Dees et al. 2025)
published by Springer (available at link.springer.com).

References

Dees, Marcus, Peter H.G. Berkhout, Claudio Di Ciccio, and Hajo A. Reijers (2025). "Decision Noise Instrument (DNI): Estimating Decision Noise in Business Processes". In: *EDOC*. Ed. by Alessandro Gianola, Renata Guizzardi, José Borbinha, Miguel Mira da Silva, and José Barateiro. Vol. 16213. Lecture Notes in Computer Science. Springer. ISBN: 978-3-032-15140-7.

BibTeX

```
@InProceedings{   Dees.etal/EDOC2025:DNI,
  author        = {Dees, Marcus and Berkhout, Peter H.G. and Di Ciccio,
                   Claudio and Reijers, Hajo A.},
  booktitle     = {EDOC},
  title         = {Decision Noise Instrument (DNI): Estimating Decision Noise
                   in Business Processes},
  year          = {2025},
  series        = {Lecture Notes in Computer Science},
  crossref      = {EDOC2025},
  keywords      = {Business Process Management; Decision analysis; Noise}
}
@Proceedings{     EDOC2025,
  title         = {Enterprise Design, Operations, and Computing - 29th
                   International Conference, {EDOC} 2025, Lisbon, Portugal,
                   September 9-12, 2025, Revised Selected Papers},
  year          = {2026},
  editor        = {Gianola, Alessandro and Guizzardi, Renata and Borbinha,
                   José and Mira da Silva, Miguel and Barateiro, José},
  isbn          = {978-3-032-15140-7},
  publisher     = {Springer},
  series        = {Lecture Notes in Computer Science},
  volume        = {16213}
}
```