Developing a Directed Graph Analysis Framework as a Method to Analyse and Present Complex Datasets: A Case Study in Tax Compliance

Zelda van der Walt[1], Johannes Petrus van der Walt[2], Marina Bornman[3] and Grietjie Verhoef[3]

[1]Faculty of Commerce, Human Sciences and Education, Namibia University of Science and Technology, Namibia

[2]Faculty of Engineering and the Built Environment, Namibia University of Science and Technology, Namibia

[3]College of Business and Economics, University of Johannesburg, South Africa

zvanderwalt@nust.na
jvanderwalt@nust.na
mbornman@uj.ac.za
gverhoef@uj.ac.za

https://doi.org/10.34190/ejbrm.22.2.3514

Abstract: Directed graphs are often used as graphical representations of interrelationships between entities. In many fields of research besides the STEM fields, datasets containing complex qualitative interrelationships are challenging to represent graphically in traditional line graphs, bar graphs, or pie charts. In addition, if quantitative data needs to be presented on top of the qualitative relationships, graphical representation becomes even more complex. As a result, datasets of this nature are often tabulated or presented in text since graphical representation is considered difficult or impractical. This paper presents a *Directed Graph Analysis Framework* that may be used to develop graphical illustrations of such complex datasets. A PhD study in employer tax compliance undertaken by the principal author is utilised as a case study in this paper (Van der Walt, Z., 2024). The framework is then used to develop a *variable interrelationship and compliance decision flow diagram* to present employer tax compliance decisions in graphical form. It is demonstrated that the method is suitable to produce a single graphical representation of a large number of variables and sub-variables, displaying the relevant qualitative and quantitative information in an easy-to-understand way. The proposed method may be applied to other fields of research where similarly complex datasets are presented.

Keywords: Directed graph, Digraph, Graphical analysis, Grounded theory, Data analysis, Tax compliance

1. Introduction

Research results are often presented in tables and graphs in fields of study such as engineering and science. A graphical presentation enables the reader to grasp the essence of the results within seconds by merely observing the trends displayed by the graphical presentation. The reader may then further observe the specific quantitative results of the research if interested in any particular values or limits. In fields of study other than STEM, such as the social sciences, it is often more challenging to analyse and present the research results in an informative and concise manner. While tables may be used to present such results, graphs are not often used since the results may not present themselves in ways readily plottable on a traditional line graph, bar graph or pie chart.

This paper presents an adaptation of directed graphs, commonly abbreviated as *digraphs,* by developing a *digraph analysis framework* (*DAF*) that not only presents relationships between datasets or variables - the typical use of digraphs - but also assists in analysing each of the variables in isolation to simplify the analysis procedure before combining the influences of all variables into a single graphical representation. The framework then enables the user to display the quantitative results of the variables in a clear and simple graphical manner. The proposed framework involves a four-step procedure where (1) the variables involved in the study are identified, (2) each variable is analysed in isolation with regard to its influences, outcomes, results or effects, (3) the combined effects of all the datasets or variables are calculated, and (4) the combined effects of all the datasets or variables are presented graphically and quantitatively to form a combined variable interrelationship diagram. The framework not only allows the researcher to display interrelationships but also to rank different interrelationships relative to quantitative importance.

Cite this article: van der Walt, Z. et al. 2024. "Developing a Directed Graph Analysis Framework as a Method to Analyse and Present Complex Datasets: A Case Study in Tax Compliance", *Electronic Journal of Business Research Methods 22(2)*, pp.13-29, https://doi.org/10.34190/ejbrm.22.2.3514

As an example of a real-world complex dataset that can be analysed using the *Digraph Analysis Framework*, the principal author conducted a tax psychology study of small to medium-sized employers' (SMEs) tax compliance concerning their employees' wages. The datasets have been slightly modified and adapted to illustrate better the development and use of the proposed *Digraph Analysis Framework*. The paper assumes the datasets without a detailed consideration of the reasons, methodology, and background regarding how the information was obtained.

2. Literature Review of Digraphs and Other Graphical Methods

A definition of a digraph is given by Metcalf and Casey (2016):

> *"A directed graph is a graph where the relationship between two vertices is a one-way relationship."*

In contrast, an undirected graph lacks the directional constructs of a directed graph. An example of a simple digraph is shown in Figure 1 below:

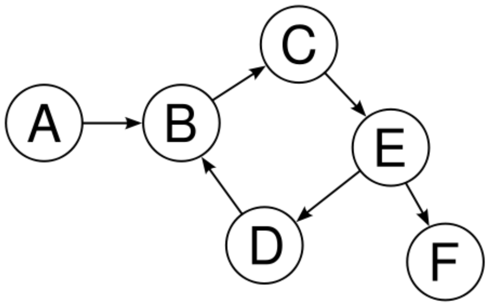

Figure 1: A simple digraph

Digraphs are graphical representations of relationships between entities, concepts, or variables. The terms *directed graph* and *digraph* appear to have been coined by Frank Harary (1955). Several variations on this theme seem to exist. *Directed Acyclic Graphs* (DAGs) have been used in many applications, including medical diagnoses (Shrier et al., 2008). DAGs represent a series of activities in such a fashion that no cyclic activity occurs. This contrasts with the cyclic activity B-C-E-D in Figure 1 above. Bang-Jensen and Gutin (2018) published a book containing many digraph types and examples. A development of digraphs presented by Gansner et al. (1993) allows for the ranking of vertices, which is closer in intent to the *digraph analysis framework* presented in this paper.

Van Rensburg (2018) employed digraphs to analyse the constraints of South African households' discretionary savings and investment habits. In her thesis, the principal author of this paper has adapted and significantly modified the digraph analysis model used by Van Rensburg to include not only relationships but also indications of the relative importance of each relationship. Furthermore, both variables and sub-variables (categories) are considered in the proposed framework.

Another related graphical method is *structural equation modelling* (SEM). SEM is often used in behavioural sciences, epidemiology (Boslaugh & McNutt 2008), business (Shelley 2006) and other fields. SEM models could use graphical representations similar to digraphs to indicate the causal connection of one phenomenon to others. Alternatively, the relationships may be represented using equations. SEM models come in two variants: covariance-based SEM (CB-SEM) and partial least squares SEM (PLS-SEM). CB-SEM is primarily used to test and confirm theories (Hair et al. 2021) and is analogous in some respects to the *digraph analysis framework* proposed in this paper, albeit significantly more complex.

A summary of the graphical methods related to digraphs considered during the study and the development of the Digraph Analysis Framework presented in this paper is given in Figure 2:

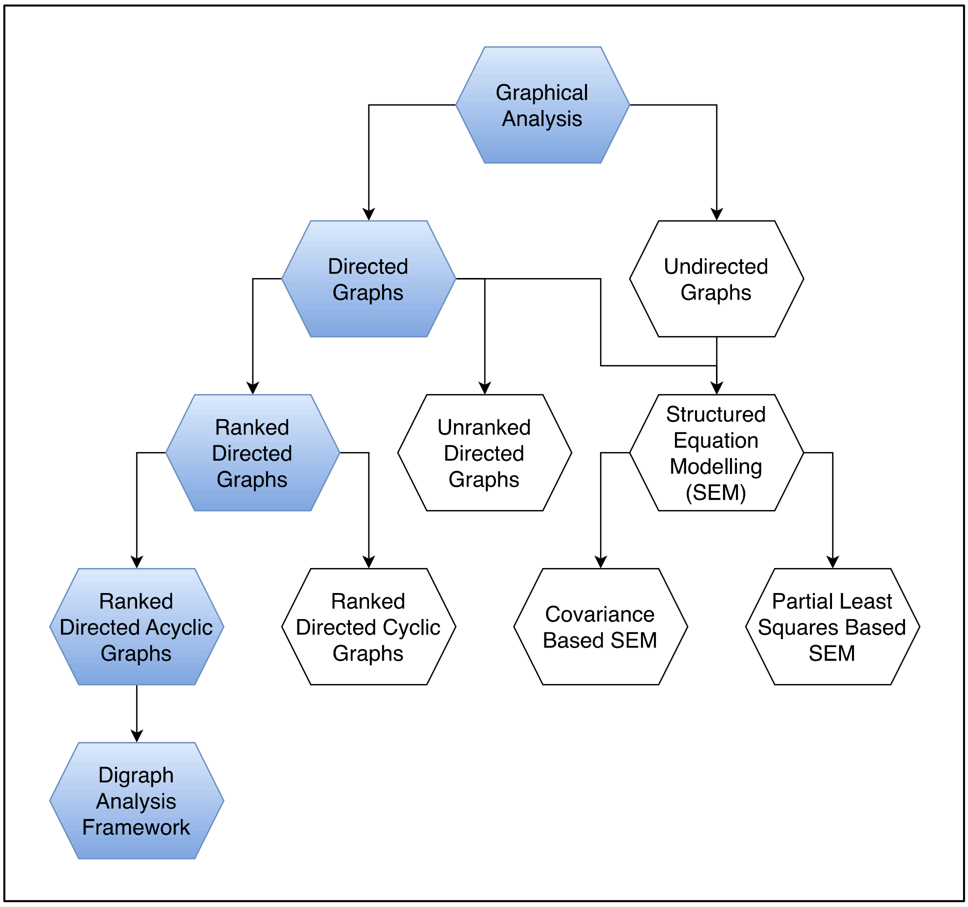

Figure 2: Graphical methods and the development of the Digraph Analysis Framework

A simple ranked acyclic digraph can thus be extended to indicate the interrelationships and decision flows between the factors identified as influencing tax compliance. Furthermore, the digraph can be adapted to assist with data analysis and to display the relative importance of the variables presented in graphical form.

3. The Data to be Analysed

The original study aimed to determine the main factors influencing employers' decisions to comply with taxation regulations regarding PAYE deductions from their employees' remuneration. A grounded theory (GT) approach was used to collect the information. In short, this approach conducted interviews with a large group of employers without a predetermined set of questions on tax compliance. As the interviews progressed, answers and comments provided by the interviewees led the discussion in new and unforeseen directions, with the variables governing their compliance decisions emerging. Three rounds of interviews were ultimately conducted.

The variables governing the compliance decisions were identified by coding the responses obtained. The number of times a specific variable was mentioned during the interviews indicated its relative importance. For example, an unfavourable view of the tax authorities was one such variable. The more frequently an interviewee mention their negative sentiments toward the tax authority as a reason for possible non-compliance, the more prominently this variable will feature in the analysis.

Seven such variables were distilled as most critical in the study. Many more variables could have been listed, but the seven identified represented the most significant contributors to the decision regarding tax compliance. Without delving into why these seven variables were recognised as the most important, this paper assumes the data as given.

Figure 3 below presents the findings of the study:

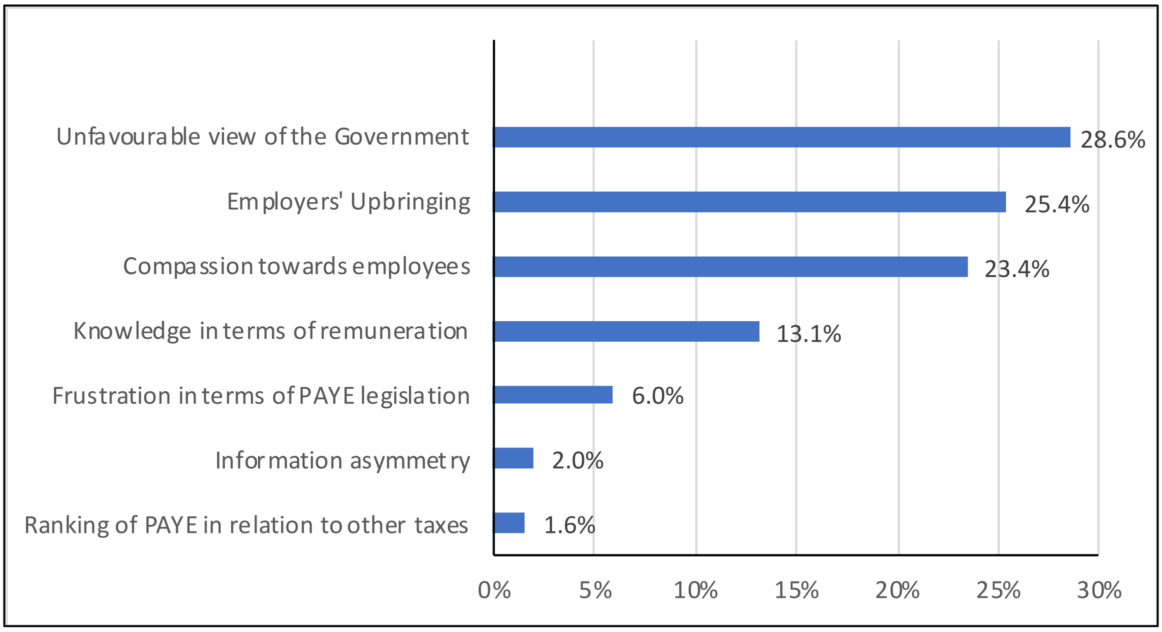

Figure 3: Compliance decision variables (in percentages)

The percentages indicate the frequency at which each variable manifested itself during the interviews relative to the other identified variables.

For clarity, the identified variables are presented in small caps in this paper. Interviewees may not necessarily mention a variable, such as an UNFAVOURABLE VIEW OF THE GOVERNMENT in precisely those terms. Instead, interviewees may remark that they perceive the tax officials as corrupt, exhibit a bad attitude towards employers, are incompetent, or cannot account for national funds, which is then assumed to be embezzled. Thus, all such expressions, called *categories* (or sub-variables, in essence), may be collected under a single *variable,* an UNFAVOURABLE VIEW OF THE GOVERNMENT in this case. This paper presents categories, or sub-variables, in italicised small caps.

Each compliance decision variable is therefore governed by one or more categories (sub-variables) underlying that variable. These categories or sub-variables explain the compliance decision in more detail. Analysing all the categories for all the variables displays the relative importance of each variable in the decision-making process. For the case study presented, the relative importance of the categories and variables (i.e., in relation to all the others) is illustrated in Figure 4 below.

Finally, with its underlying categories, each variable results in one of three outcomes: NON-COMPLIANCE, ENFORCED COMPLIANCE, or VOLUNTARY COMPLIANCE. The data analysis needs to show how each variable and its categories influence the employer's decision to either comply or not comply with the tax legislation.

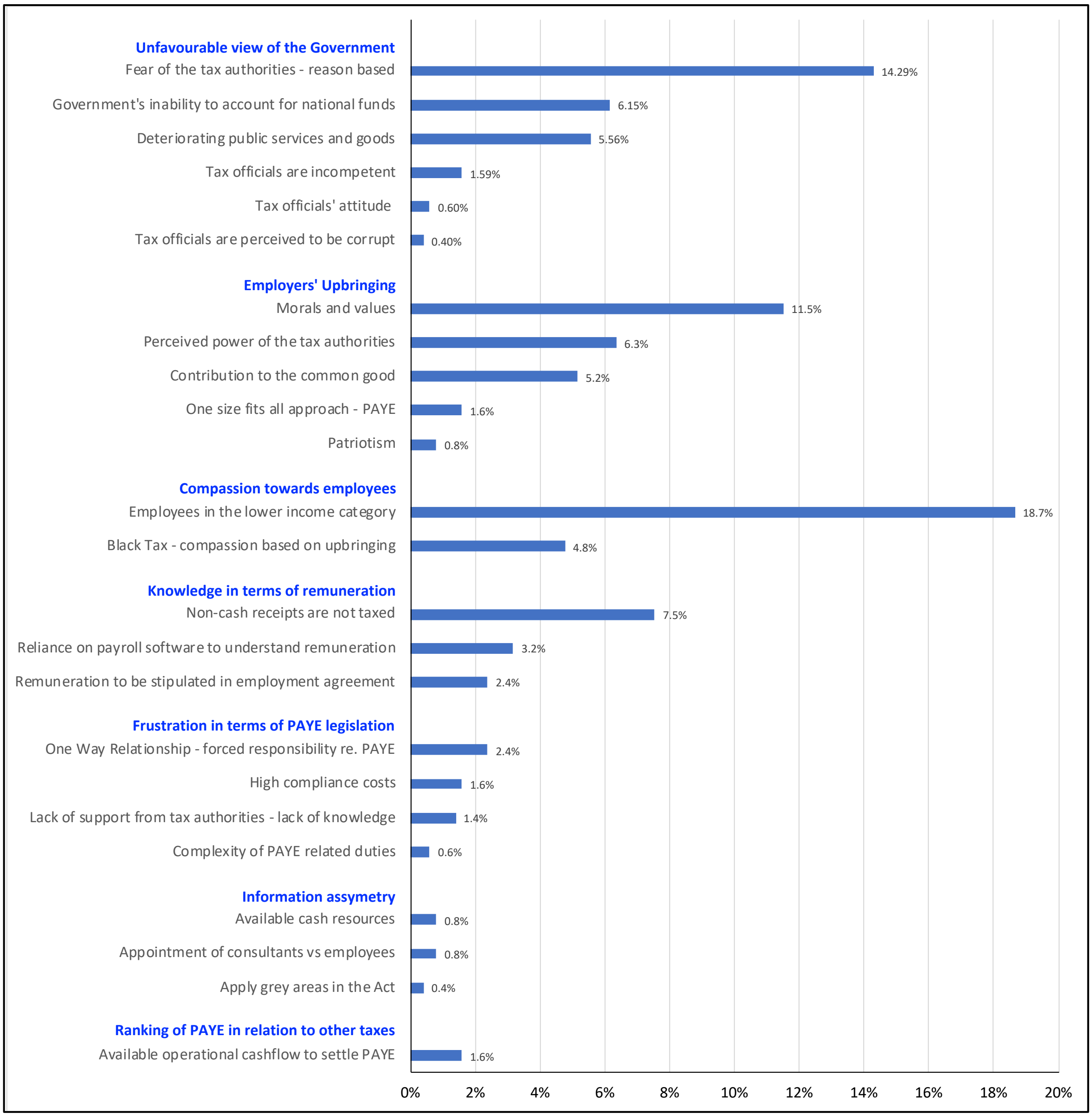

Figure 4: Compliance decision variables and categories (in percentages)

4. Development of the Variable Interrelationship and Decision Flow Digraph Using the Digraph Analysis Framework

To establish the interrelationships and decision flows between the seven identified decision-making variables, each variable and its underlying categories were individually considered in relation to each of the other variables. These interrelationships then determined the flow of decisions in the compliance or non-compliance options of the interviewed employers.

To provide the reader with maximum clarity, each variable and its associated categories are analysed one at a time in this section, each under its own subheading. Although this results in a rather long Section 4, it provides a clear progression of the analysis process of all the variables, first from simple variables, then progressing to quite complex ones. The graphical analyses of each of the seven variables are presented in Sections 4.1 to 4.7.

Furthermore, the analysis is divided into four distinct steps or procedures spanning across Section 4 and Section 5:

- Step One: Listing of all variables (Section 4 below).
- Step Two: Analysis of the influence of each variable (one at a time) on all other variables (Sections 4.1 to 4.7).

- Step Three: Calculation of the combined influences of all the variables considered (Section 5).
- Step Four: Graphical presentation of the combined influences of all variables considered (Section 5).

Using the *DAF*, the **first step** of the procedure is to create a graphical presentation listing all the variables and outcomes, as shown in Figure 5 below. This case study has seven variables and three outcomes.

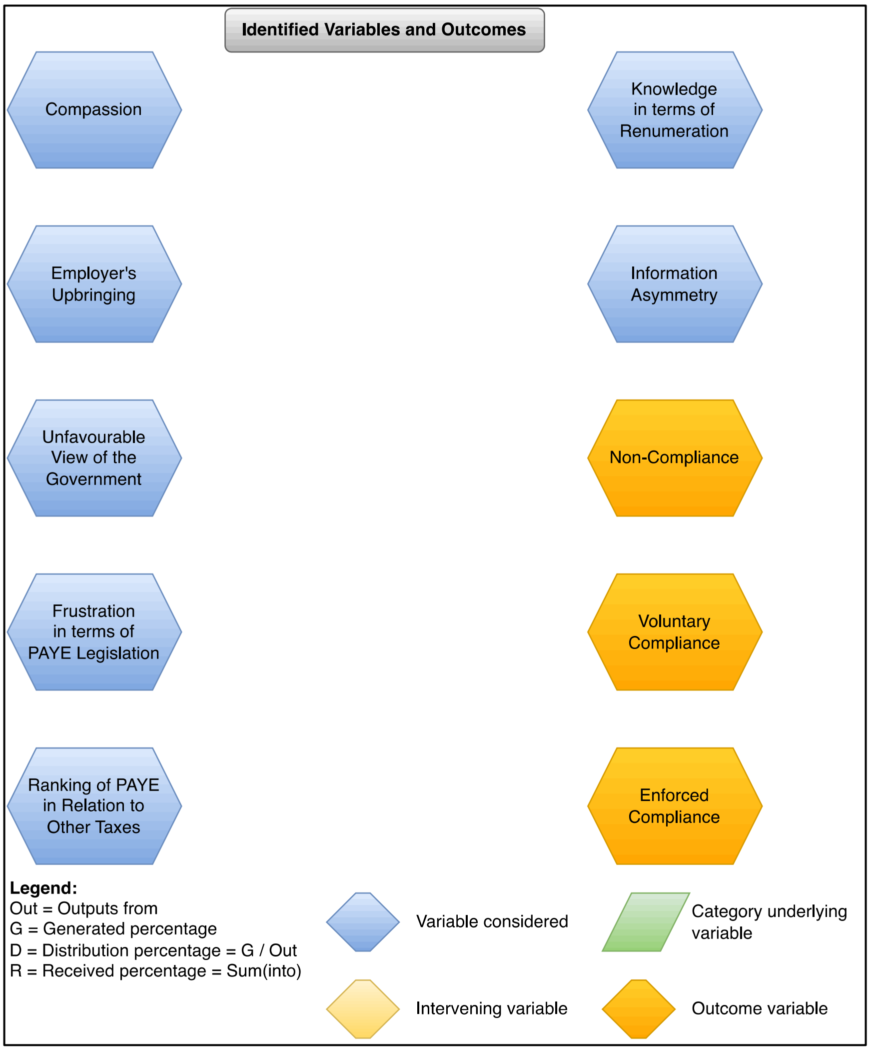

Figure 5: Placing the variables and outcomes on an empty digraph

In the **second step** of the *DAF* procedure, each variable, with its underlying categories, is plotted against the appropriate outcome or intervening variables. This is done one variable at a time to carefully calculate and assign the generated frequencies as percentages, and distribution decisions as percentages and receipt percentages. To demonstrate this, we will first analyse a straightforward variable, for example, the employer's ranking of the importance of PAYE to other taxes such as VAT, and so on.

Figure 6 illustrates the effect of the perceived importance of the PAYE legislation relative to other taxes (the variable being considered) which the employer must deal with. Suppose an employer concludes that PAYE is not of primary importance and that this employer happens to be experiencing operational cash flow constraints. Such an employer may be tempted to explore non-compliance opportunities and ultimately may not comply with regulations.

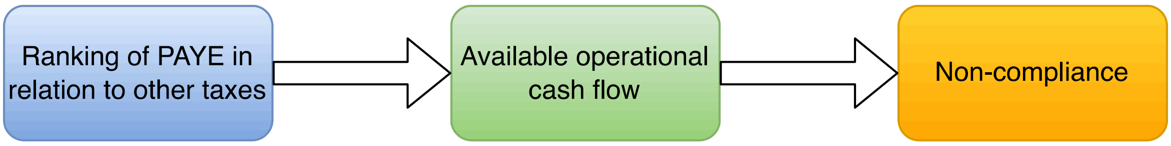

Figure 6: Influence of perceived PAYE ranking on compliance

In the following sections, each identified variable, as shown in Figure 5 above, with its associated categories, will be plotted and analysed relative to all the other variables.

4.1 RANKING OF PAYE IN RELATION TO OTHER TAXES as a variable in the compliance decision

From Figure 4, the variable RANKING OF PAYE IN RELATION TO OTHER TAXES had an interview occurrence frequency (i.e., the number of times this variable was activated during the interview via its associated categories) of 1.6%. The interview occurrence frequency is displayed as the *Generated frequency percentage* (G) on the digraph shown in Figure 7 below for this variable.

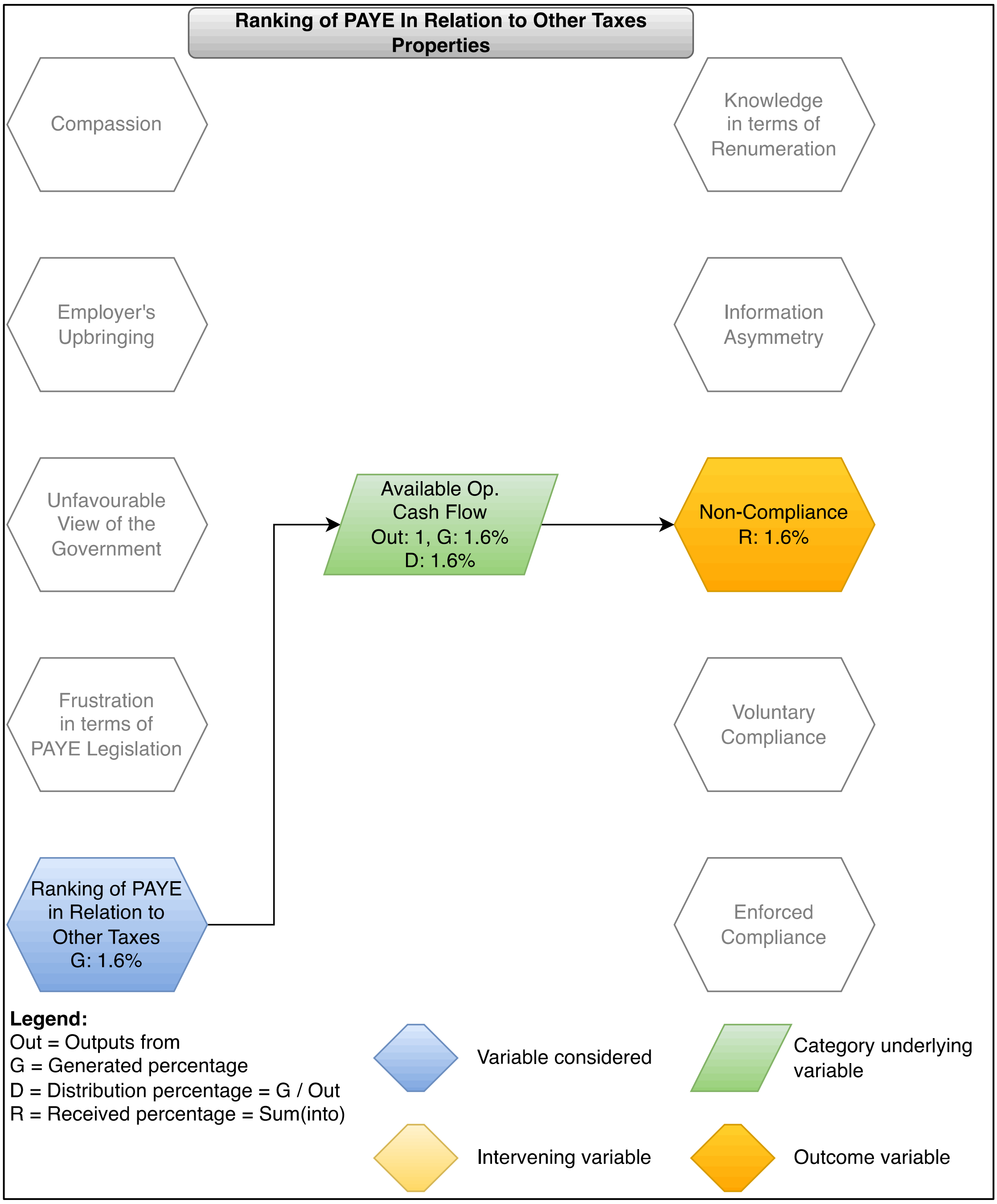

Figure 7: The RANKING OF PAYE IN RELATION TO OTHER TAXES as a variable in the compliance decision

The variable under consideration, RANKING OF PAYE IN RELATION TO OTHER TAXES, exists because employers' comment on the option not to comply with tax regulations due to a lack of available operational cash flow in the company. Hence, the category *AVAILABILITY OF OPERATIONAL CASH FLOW* represents 1.6% of the total compliance or non-compliance decision.

The figure shows that this category generated (G) 1.6% of the compliance decision. Since this category is the only one linked to the associated variable, this variable is therefore also assigned a generated percentage (G) of 1.6%. Lastly, since the category under discussion also has only one output to a single recipient variable, this recipient variable is allocated a received decision (R) percentage of 1.6%.

However, variables are often associated with several categories and more than one recipient variable. We discuss such cases below.

To be clear, the categories, and not the variables, generate the frequency percentages. The sum of the generated percentages is then assigned to the associated variable and distributed to the receiving variables, as will be shown in subsequent variable analyses.

4.2 EMPLOYERS' FRUSTRATION WITH PAYE LEGISLATION as a Variable in the Compliance Decision

From Figure 4 above, the variable FRUSTRATION WITH PAYE LEGISLATION has a generated frequency percentage (G) of 6.0%. This sum-total is comprised of four identified categories as shown in Figure 8 below: *ONE-WAY RELATIONSHIP* between the tax authorities and employers in their role as tax agents (2.4%); *HIGH COMPLIANCE COSTS* (1.6%); *LACK OF SUPPORT FROM THE TAX AUTHORITIES* (1.4%); and the *COMPLEXITY OF PAYE-RELATED DUTIES* (0.6%).

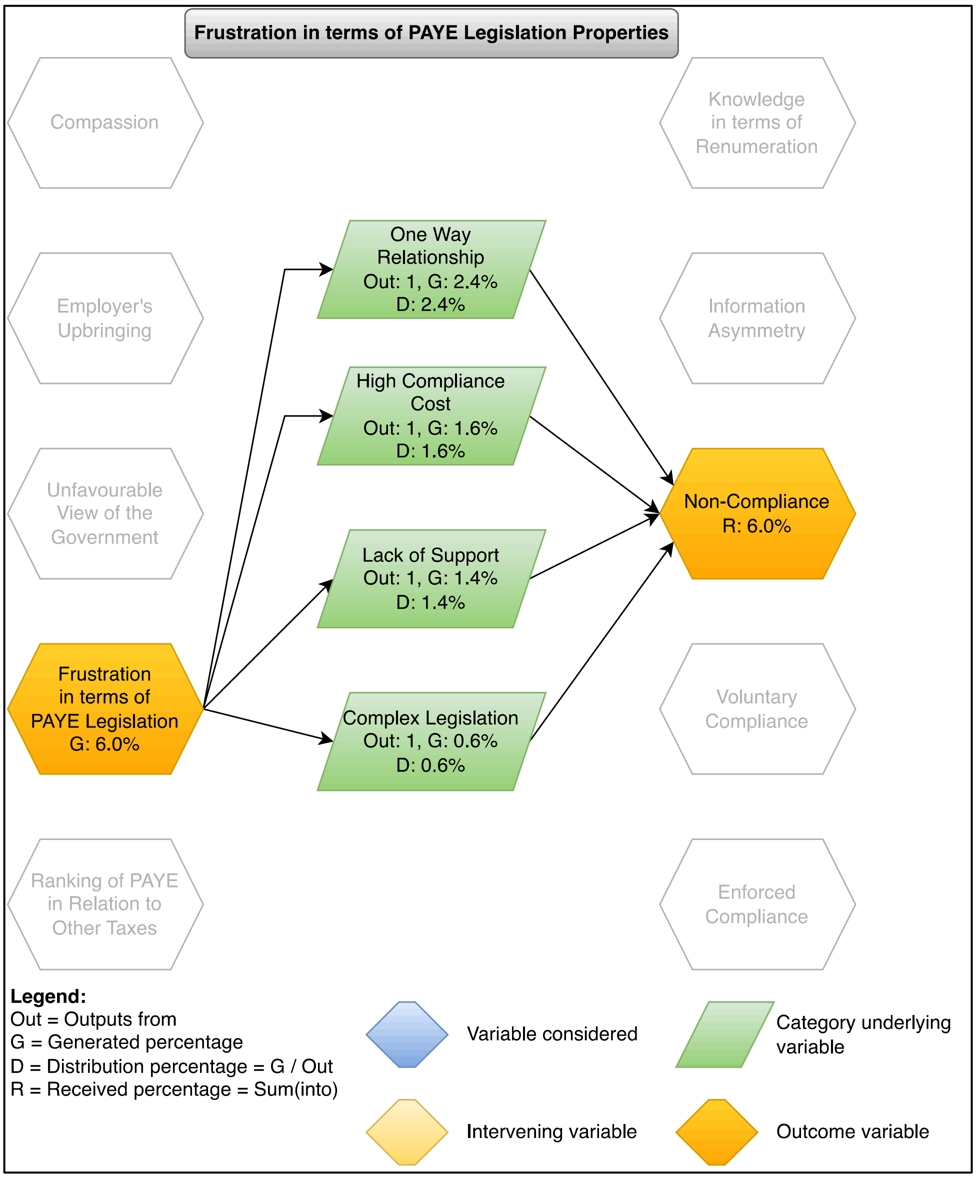

Figure 8: EMPLOYERS' FRUSTRATION WITH PAYE LEGISLATION as a variable in the compliance decision

During the interviews, the category *ONE-WAY RELATIONSHIP* was cited 2.4% of the time as a reason for non-compliance. Thus, that category generated (G) 2.4% of the frustration the employer mentioned. Since this category has only one link (Out: 1) to the outcome variable NON-COMPLIANCE, it *distributes* the complete 2.4% (D: 2.4%) to the recipient variable.

In this simple example, the sum of the generated frustrations of the four categories equals the 6% indicated in the considered variable on the left-hand side. Furthermore, since all four categories contribute to the same outcome variable, NON-COMPLIANCE on the right-hand side, its *received* percentage (R) also equals the total of the categories.

An example of a slightly more complex variable, the EMPLOYER'S UPBRINGING, is discussed next.

4.3 EMPLOYERS' UPBRINGING as a Variable in the Compliance Decision

As shown in Figure 4, the variable under consideration, the EMPLOYER'S UPBRINGING*,* consists of five categories, as depicted in Figure 9 below. It had a generated frequency percentage (G) of 25.4% and is comprised of the categories *MORALS AND VALUES* (11.5%), *CONTRIBUTION TO THE COMMON GOOD* (5.2%), *PATRIOTISM* (0.8%), *ONE-SIZE-FITS-ALL APPROACH IS INAPPROPRIATE* (1.6%), and *FEAR OF THE TAX AUTHORITIES* (6.3%).

This variable and its categories are different from the variables discussed above in terms of the output distribution of the categories. Each category still has only one output, therefore the generated percentage (G) and the distributed percentage (D) in each category are equal in all cases. The difference, however, is that the five categories do not all contribute to the same outcome variable. Each outcome variable receives a percentage from one or more categories. These percentages are then summed and reflected as the received percentage (R), as shown in Figure 9.

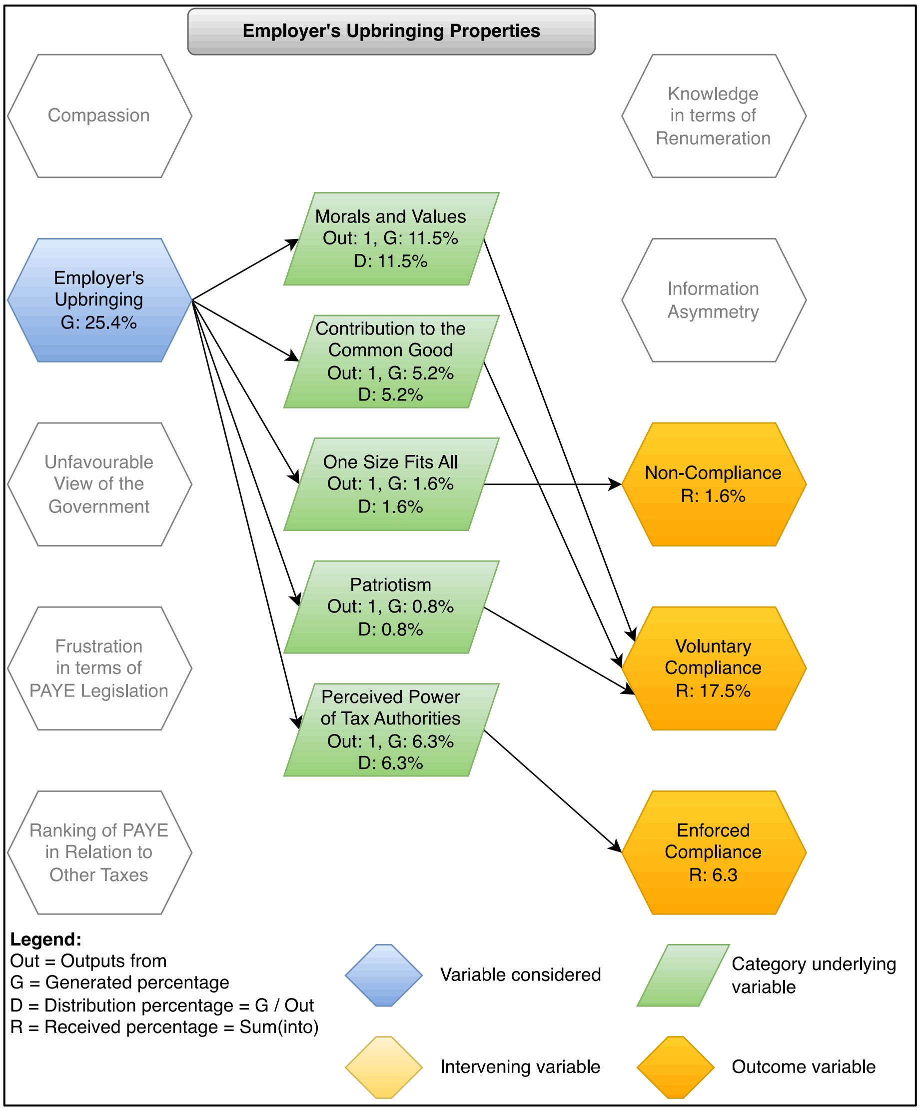

Figure 9: UPBRINGING as a variable in the compliance decision

4.4 INFORMATION ASYMMETRY as a Variable in the Compliance Decision

In Figure 4 above, the variable INFORMATION ASYMMETRY generated a frequency percentage (G) of 2.0%. The variable is also described by the categories of *AVAILABILITY OF CASH RESOURCES* (0.8%), *THE APPOINTMENT OF CONSULTANTS VERSUS EMPLOYEES* (0.8%), and *THE APPLICATION OF GREY AREAS IN THE ACT* (0.4%).

This variable and its associated categories are straightforward, and the related digraph analysis is shown in Figure 10 below. Information asymmetry is when the employer, as a tax agent of the tax authorities, does not share all information with the tax authorities.

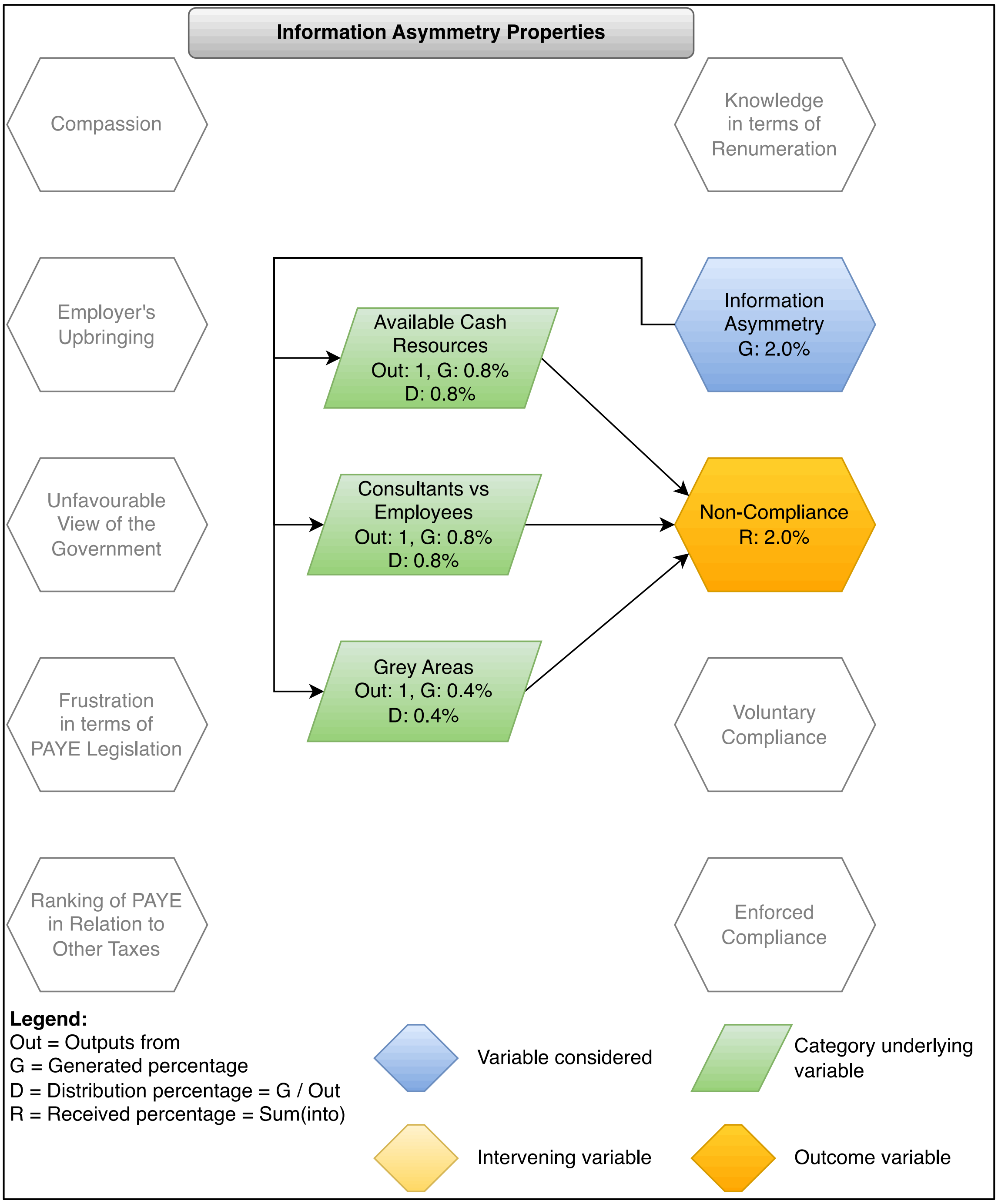

Figure 10: INFORMATION ASYMMETRY as a variable in the compliance decision

4.5 EMPLOYERS' COMPASSION FOR EMPLOYEES as a Variable in the Compliance Decision

From Figure 4 above, the variable EMPLOYERS' COMPASSION FOR EMPLOYEES generated a frequency percentage (G) of 23.5%. The variable is described by the categories: *COMPASSION FOR EMPLOYEES IN THE LOWER INCOME CATEGORIES* (18.7%) and *BLACK TAX* (4.8%), as shown in Figure 11 below.

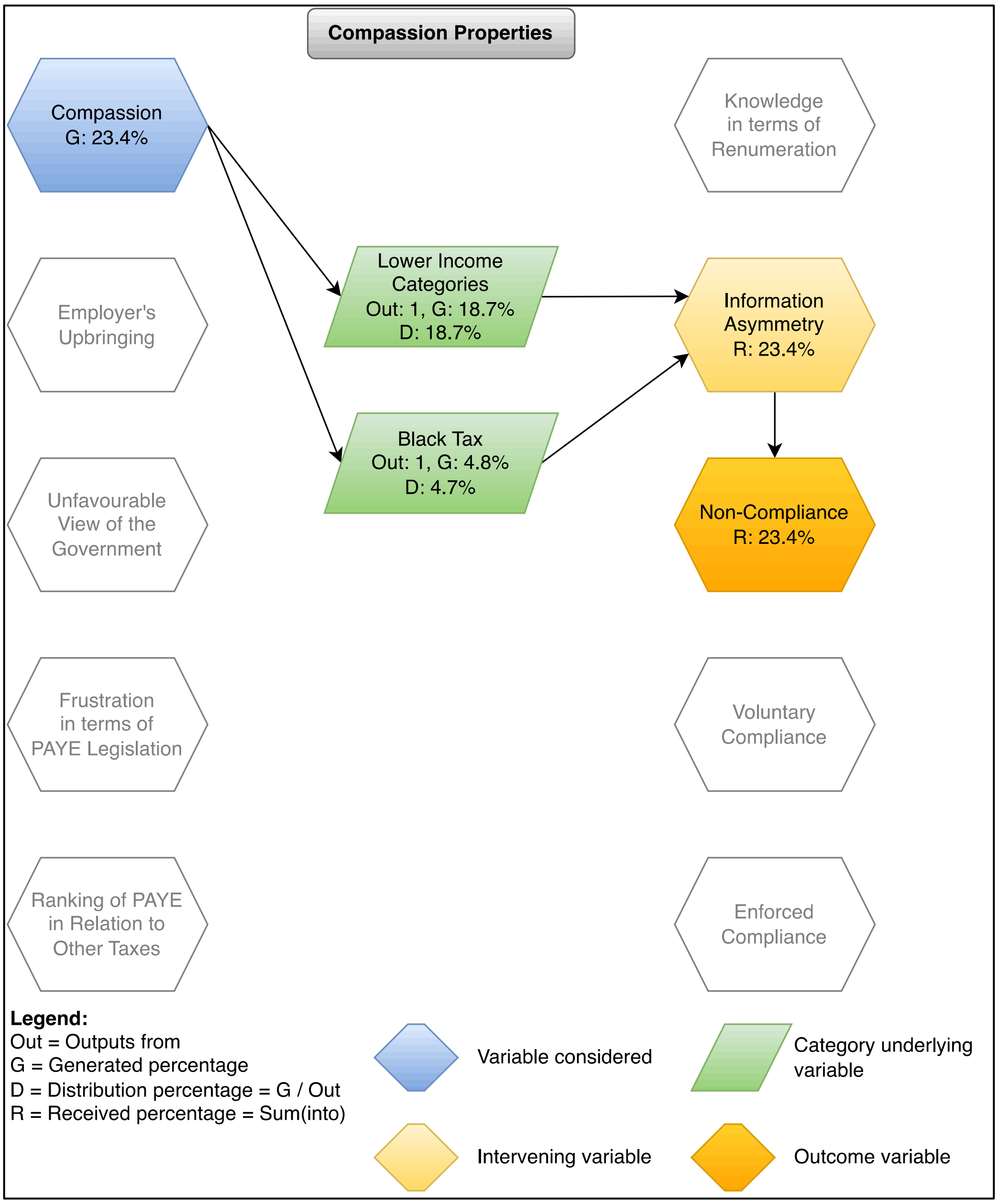

Figure 11: COMPASSION FOR EMPLOYEES as a variable - interrelationships and decision flows

However, both categories cause an INFORMATION ASYMMETRY between the employer and the tax authorities since the tax authorities are not informed about the compassionate behaviour of the employer who pays the taxi fares of a low-income worker in cash, for example. In the case of the COMPASSION variable, therefore, the two associated categories result in non-compliance via INFORMATION ASYMMETRY. In this example, INFORMATION ASYMMETRY is an *intervening variable* that generates no compliance decision percentage. Instead, it combines and transfers the decision percentages of the preceding categories to the recipient variable, NON-COMPLIANCE, as shown in Figure 11.

4.6 UNFAVOURABLE VIEW OF THE GOVERNMENT as a Variable in the Compliance Decision

From Figure 4 above, the variable UNFAVOURABLE VIEW OF THE GOVERNMENT generated a frequency percentage (G) of 30.2%. The variable is comprised of the categories: *TAX OFFICIALS ARE INCOMPETENT* (1.6%), *TAX OFFICIALS' ATTITUDE*

(0.6%), *TAX OFFICIALS ARE PERCEIVED TO BE CORRUPT* (0.4%), *DETERIORATING PUBLIC SERVICES AND GOODS* (5.6%), *INABILITY TO ACCOUNT FOR NATIONAL FUNDS* (6.1%) and *FEAR OF THE TAX AUTHORITIES* (14.3%).

Each underlying category has one connection with the variable under consideration, to which it assigns the generated percentage (G). However, each category may contain multiple outputs to different recipient variables (R).

As shown in Figure 12 below, only one category, *FEAR OF TAX AUTHORITIES*, directly results in a compliance decision (ENFORCED COMPLIANCE). The remaining categories contribute to intervening variables, all analysed as independent variables in the preceding paragraphs. Variables such as COMPASSION therefore act as *driver variables* and as *intervening variables*. Driver variables cause decisions to be taken and thus generate a decision percentage. Intervening variables collect the generated decisions of driver variables and then distribute them to outcome variables.

Figure 12 shows that the category of *INCOMPETENT OFFICIALS* contributes to the intervening variables of INFORMATION ASYMMETRY and FRUSTRATION IN TERMS OF PAYE LEGISLATION. This category has a generated decision percentage (G) of 1.6%. Due to a lack of more detailed low-level information in this study, it was assumed that an equal split of decision percentages (0.8%) would flow to each intervening variable. If more research had been conducted on the specific importance of each split, different percentages could have been assigned to each intervening variable.

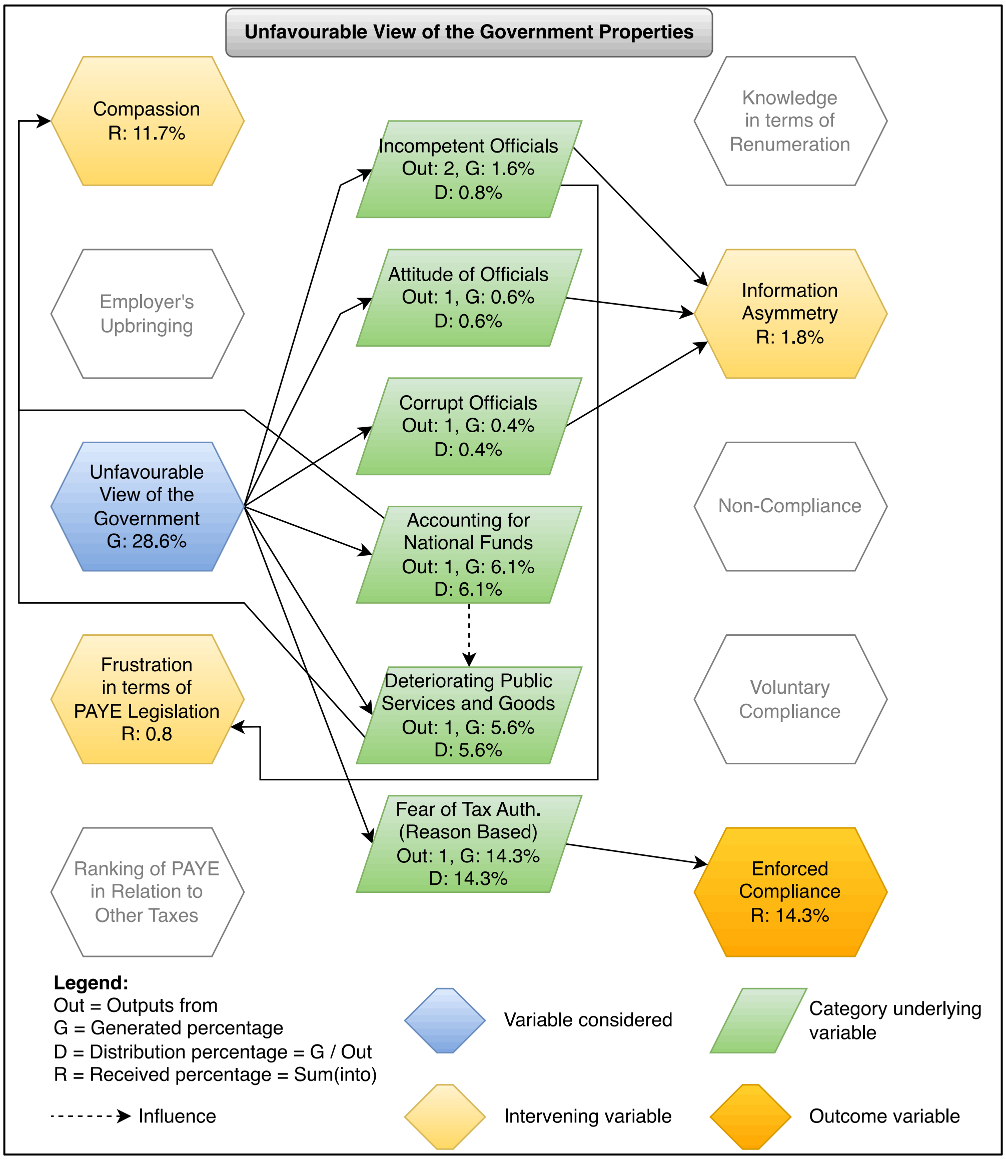

Figure 12: UNFAVOURABLE VIEW OF THE GOVERNMENT as a variable: interrelationships and decision flows

4.7 KNOWLEDGE OF REMUNERATION as a Variable in the Compliance Decision

From Figure 4 above, the variable KNOWLEDGE OF REMUNERATION generated a frequency percentage (G) of 13.1%. The categories that further describe the variable are *NON-CASH RECEIPTS* (7.5%), *RELIANCE ON PAYROLL SOFTWARE AS A MEANS TO ENHANCE KNOWLEDGE* (3.3%) and *VIEW THAT REMUNERATION SHOULD BE STIPULATED IN THE EMPLOYMENT AGREEMENT* (2.3%).

These straightforward results are shown in the associated digraph in Figure 13 below.

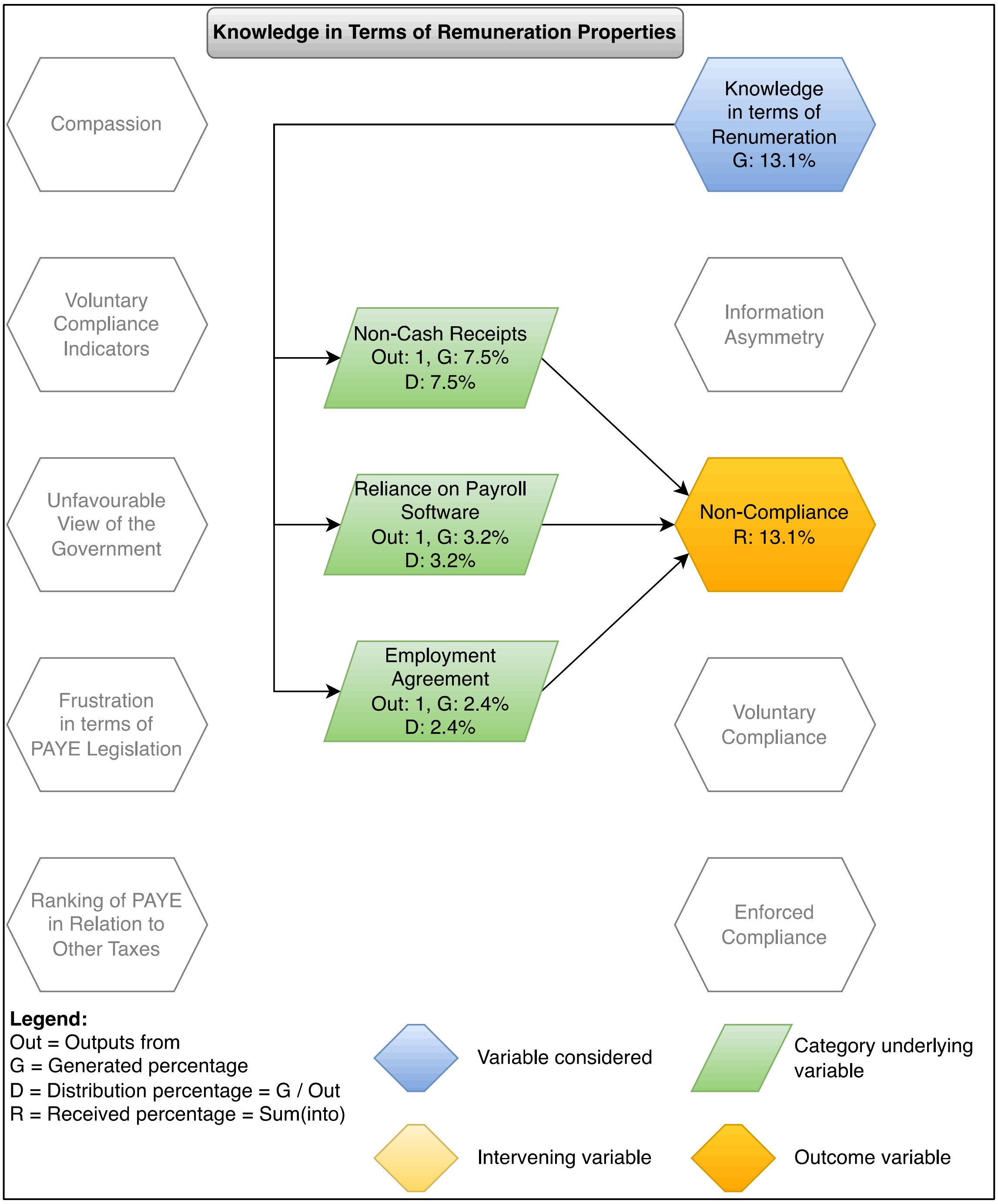

Figure 13: KNOWLEDGE OF REMUNERATION as a variable in the compliance decision

5. Results Assembly

The **third step** of the *DAF* procedure is to calculate the combined influence of each variable on all the other variables. This calculation was performed in a spreadsheet, and its outcomes are presented in Figure 14 below. As analysed from Figure 7 to Figure 13, the variables are listed on the left (*Variables Generating Influence*). The intervening and outcome variables receiving influence are listed in the top row. The influence of each variable in the column on the left on each of the other variables can be seen by following its row horizontally to the right to the cell where the two variables concerned intersect. For example, Compassion has an influence of 23.4% on Information asymmetry. The right-most column reflects the total influence each of the influence-generating variables in the left-most column has on all receiving variables, and this total is equal to the generated percentage (G) of the variable on the left. The first grey row at the bottom (Percentages Received - R) shows the inputs into each intervening or outcome variable.

Figure 11 shows that Compassion results in Non-Compliance via the intervening variable *Information Asymmetry*. This result is also reflected in Figure 14 below. Similarly, the variable Unfavourable view of the government contributes to both Information Asymmetry and Compassion. Since Information Asymmetry leads to Non-Compliance, the generated percentages 11.7% and 25.2% are combined (36.9%) in the second grey row at the bottom of the table.

The third grey row combines all the generated percentages G into the three outcome variables. Since Information Asymmetry (36.9%) only leads to Non-Compliance, Information Asymmetry (36.9%) and Non-Compliance generated from other variables (25.1%) are combined to result in the total Non-Compliance output variable (62.0%).

	Intervening Variables							***Outcome Variables***			
Variables Generating Influence	Ranking of PAYE re Other Taxes	Frustration re PAYE Legislation	Employer's Upbringing	Unfavourable View of the Government	Compassion	Information Asymmetry	Knowledge re Remuneration	Non-Compliance	Voluntary Compliance	Enforced Compliance	**Percentages Generated (G-values)**
Ranking of PAYE re Other Taxes								1.6			1.6
Frustration re PAYE Legislation								6.8			6.8
Employer's Upbringing								1.6	17.5	6.3	25.4
Unfavourable View of the Government					11.7	1.8				14.3	27.8
Compassion						23.4					23.4
Information Asymmetry								2.0			2.0
Knowledge re Remuneration								13.1			13.1
Percentages Received (R-values)					11.7	25.2		25.1	17.5	20.6	100
Combining Receiving Variables (R-values)						36.9		25.1			
Total Compliance Outcome Values (O-values)								62.0	17.5	20.6	100

Figure 14: Variables and Outcomes Results

The **fourth and final step** of the *DAF* procedure is to combine the digraphs developed from Figures 7 to 13 and the total percentages calculated in Figure 14 into a single *Variable interrelationship and decision flow digraph*. This digraph shows not only the relationships between all the variables but also the flow of the compliance decision-making process, thus resulting in an estimate of percentages for the *outcome variables*: Non-compliance, Voluntary compliance, and Enforced compliance. This combined *Variable interrelationship and decision flow digraph* is shown in Figure 15

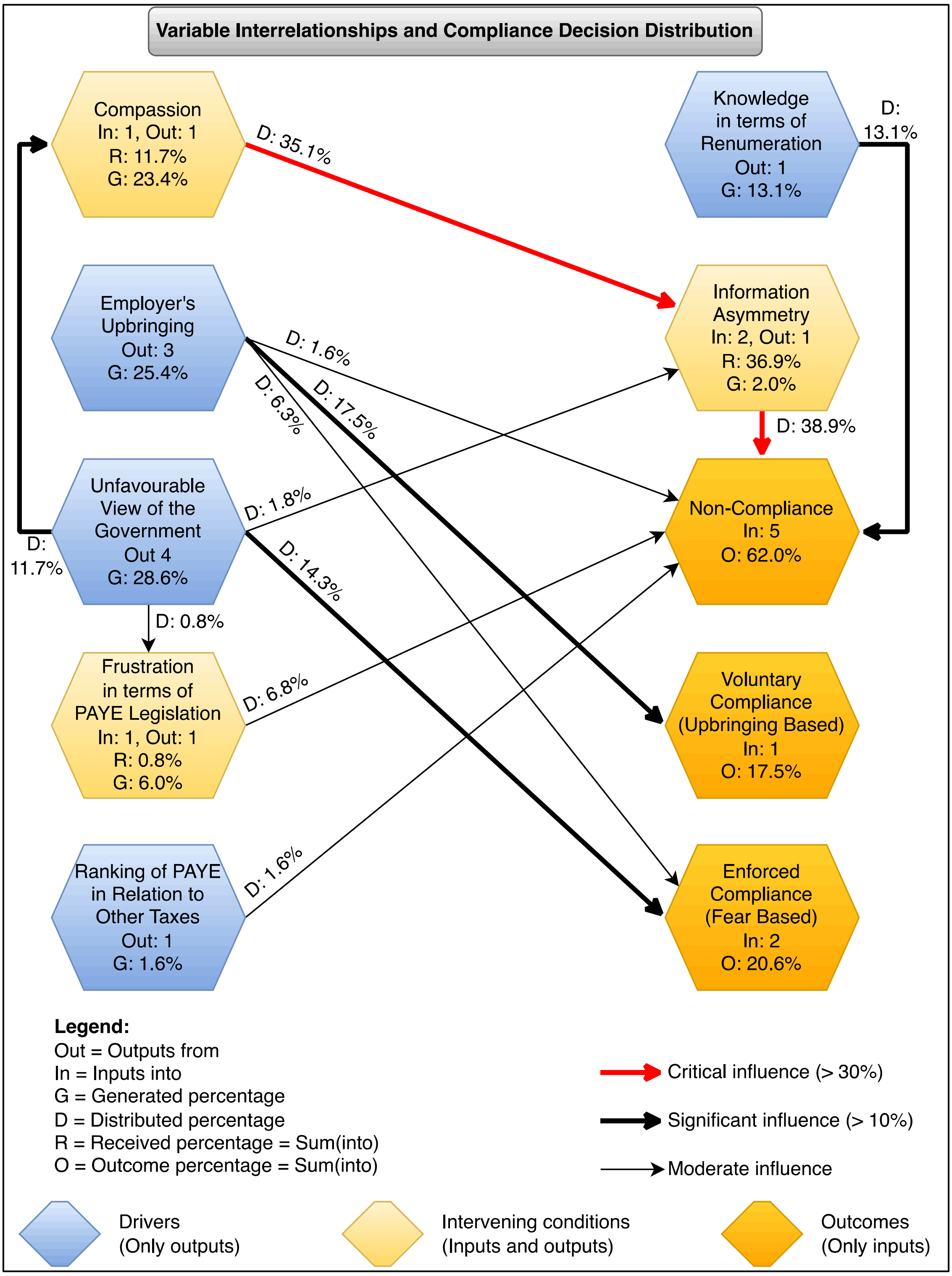

Figure 15: Variable interrelationships and compliance decision flow diagram

Variables with no inputs, i.e., variables not influenced by other variables, are *Drivers*. On the other hand, variables that only have inputs are *Outcomes*. Variables that have both inputs and outputs are *Intervening variables*.

The G-values shown in the variables in Figure 15 indicate the percentage of the occurrence (generated) frequency during the interviews. Similarly, the R-values represent the sum of all the G-values received into the particular recipient variable. The label *In* indicates the number of transfers into a variable, and the label *Out* shows the number of transfers out of the variable. At the start of each transfer arrow, the numeric value indicates the percentages transferred or distributed (D-value) to the recipient variable. Recipient variables show the sum of all received percentages as R-values. The sums of the inputs into the *Outcomes* are shown as O-values.

In the V*ariable interrelationship and decision flow digraph*, Figure 15, red arrows indicate critical drivers exceeding a specific limit (30% selected in this case). Thick black arrows indicate significant drivers in the decision-making process (exceeding 10% in this case). Standard black arrow lines denote moderate influences (10% and less). The limits selected for these levels of influence are arbitrary and are set by the researcher to best reflect the data being analysed.

6. Data Analysis

Analysing the data presented in the final *Variable interrelationship and decision flow digraph,* as shown in Figure 15, immediately reveals the following significant trends:

- INFORMATION ASYMMETRY contributes to 38.9% of non-compliance decisions, as indicated by a thick red arrow.
- COMPASSION considerations alone contribute to 35.1% of the decisions to exploit non-compliance opportunities, also indicated by a thick red arrow.
- The UNFAVOURABLE VIEW OF THE GOVERNMENT variable increases compassionate action by a significant 11.7%, indicated by a thick black arrow.
- An EMPLOYER'S UPBRINGING contributes 17.5% to VOLUNTARY COMPLIANCE, also as indicated by a thick black arrow.
- An UNFAVOURABLE VIEW OF THE GOVERNMENT (fear of the tax authorities' power) may contribute to 14.3% of ENFORCED COMPLIANCE (thick black arrow).
- The contribution of LACK OF KNOWLEDGE, i.e., employers not complying with the tax law simply because they misunderstand the law, is 13.1%, as shown by the last thick black arrow.
- All the other influences may be considered moderate or minor (thin black arrows).
- Finally, the *Outcomes* are illuminating. They indicate that respondents are 62% likely not to comply, with VOLUNTARY COMPLIANCE at only 17.5% and ENFORCED COMPLIANCE similarly low at 20.6%.

If the tax authorities wanted to improve compliance, they would be well advised to focus on resolving the critical and significant influences resulting in non-compliance.

7. Conclusion

This paper presented a *Directed Graph (digraph) Analysis Framework* (*DAF*), which may be used to analyse and graphically display complex datasets that include both qualitative and quantitative information and cannot easily be presented using traditional graphing methods. In the STEM fields, simple line graphs, more advanced 3-D surface plots, radar plots, bar graphs and pie charts may suffice; however, in many other fields of research, such as the humanities and other fields such as the tax psychology case study presented in this paper, none of these traditional graphing methods would suffice. Numerous other fields of research outside the traditional STEM fields, such as the humanities, may benefit from this approach, especially where a combination of qualitative and quantitative data needs to be presented graphically.

A four-step method was presented that allows several interdependent variables, each with associated sub-variables (categories), to be analysed against all other variables and plotted graphically.

The four-step method is summarised as follows:

Step 1: Create a graphical presentation listing all the variables and outcomes.

Step 2: Plot each variable, with its underlying categories or sub-variables, against the appropriate outcome or intervening variables.

Step 3: Calculate the combined influence of each variable on all the other variables.

Step 4: Combine all the digraphs developed in Step 2 and the total percentages calculated in Step 3 into a single variable interrelationship and decision flow digraph.

The final *Variable interrelationship and decision flow digraph,* as shown in Figure 15, displays the information obtained in the case study. The interrelationships are clearly demonstrated using arrow lines. The level of importance of these relationships is visually presented in terms of the colours and thicknesses of the arrow lines. The three variable types used in this case study, *Driver*, *Intervening* and *Outcome*, are visually distinguished by using different colours. Should the reader need a more precise data resolution, the percentages generated for each interrelationship are also shown on the arrow lines.

This case study presented the analysis of seven variables and twenty-four sub-variables. The influence of each of the variables on all others was analysed using a simple procedure and combined into a relatively simple and understandable graphical representation, as shown in Figure 15. It is unlikely that a traditional graphing solution could be used to present the qualitative and quantitative interrelationships of a total of thirty-one variables and sub-variables in the clear and concise manner as presented in this paper.

The *Digraph Analysis Framework* (*DAF*) may be modified and extended to suit many other types of datasets that researchers may encounter.

References

Bang-Jensen, J. and Gutin, G., 2018. *Classes of directed graphs*. Springer Monographs in Mathematics, Springer. https://www.springer.com/series/3733

Boslaugh, S. and McNutt, L.A., 2008. *Structural Equation Modeling*. Encyclopedia of Epidemiology. doi 10.4135/9781412953948.n443, ISBN 978-1-4129-2816-8.

Brassard, M. and Ritter, D., 2010. *The memory jogger 2: Tools for continuous improvement and effective planning*. (2nd Ed). Goal / QPC.

Gansner, E.R., Koutsofios, E., North, S.C., and Vo, K. -P., 1993. *A technique for drawing directed graphs*. IEEE Transactions on Software Engineering, vol. 19, no. 3, pp. 214-230, March 1993, doi: 10.1109/32.221135.

Hair, J.F. et. al., 2021. *An Introduction to Structural Equation Modeling*. In: Partial Least Squares Structural Equation Modeling (PLS-SEM) Using R. Classroom Companion: Business. Springer, Cham. https://doi.org/10.1007/978-3-030-80519-7_1

Harary, F., 1955. *The number of linear, directed, rooted, and connected graphs*. Transactions of the American Mathematical Society, Vol. 78, No. 2 (Mar. 1955), pp. 445-463 (19 pages). Published By: American Mathematical Society.

Metcalf, L. and Casey, W., 2016. *Cybersecurity and Applied Mathematics*. ISBN 978-0-12-804452-0. Elsevier 2016.

Metcalf, L. and Casey, W., 2016. *Graph theory*. Cybersecurity and Applied Mathematics. Book, Ch 5. https://www.sciencedirect.com/book/9780128044520/cybersecurity-and-applied-mathematics#book-info

Shelley, M.C. (2006). *Structural Equation Modeling*. Encyclopedia of Educational Leadership and Administration. doi 10.4135/9781412939584.n544, ISBN 978-0-7619-3087-7.

Shrier, I. and Platt, R.W., 2008. *Reducing bias through directed acyclic graphs*. BMC Medical Research Methodology 8, 70 (2008). https://doi.org/10.1186/1471-2288-8-70

Van der Walt, Z., 2024. *Compliance behaviour of employers towards employee taxes: a case of Namibian MSMEs*. PhD Thesis. University of Johannesburg.

Van Rensburg, E., 2018. *Constraints to discretionary savings and investment in South Africa*. PhD Thesis, Ch. 3, University of Johannesburg.