

Subjective Interpretation of Directive Terms: Methodology and Preliminary Results

Christopher K. Frantz*, Swetha Keertipati†, Martin K. Purvis† and Maryam Purvis†

* College of Enterprise and Development, Otago Polytechnic, New Zealand

E-mail: christopher.frantz@op.ac.nz

† Department of Information Science, University of Otago, New Zealand

E-mail: keesw879@student.otago.ac.nz, {martin.purvis,maryam.purvis}@otago.ac.nz

Abstract—This work attempts to capture linguistic aspects of norms, such as the ‘musts’ and ‘must nots’ that govern our daily life, empirically and provide the foundation for a general classification of terms based on their prescriptiveness. The communication of obligations and prohibitions among humans is generally unproblematic and contextualised by a common cultural background; the declaration of *musts* and *must nots* is largely interpreted unambiguously. The boundary cases between those extremes, however, have hardly been explicitly explored. Consequently, this work concentrates on the wider spectrum of prescriptions and attempts to shed light on the boundary cases that lie between the extremal obligations and prohibitions. We propose a methodology to attempt a systematic empirical grounding of directives in human language. Focusing on the English language as the most important language for cross-cultural discourse, we attempt to identify individuals’ understandings for specific terms in order to determine whether a generalisable understanding of prescriptiveness can be achieved. Though this paper is primarily centred on the methodological perspective, we provide preliminary results from an initial study that informs methodological refinements.

Keywords—Cross-cultural Communication, Norms, Directives, Prescriptiveness, Classification, Survey, Methodology, Dynamic Deontics

I. INTRODUCTION

Communication is the cornerstone of coordination and cooperation in human societies, an aspect that is evermore important in modern multi-cultural societies shaped by different backgrounds, diverse social structures, and manifestations of social behaviour (institutions). Whether expressed as conventions of behaviour, unwritten social norms [1] that individuals abide to, or as formal legal prescriptions [2],¹ a unified understanding requires communication, be it by imitation, observation, in oral or written form.

This poses the question in how far human understanding is unified and unambiguous, a challenge that is constitutive for the terminological specifications entertained in legal disciplines [4]. Since legal prescriptions, or rules, are aligned with categorisations into *rights*, *obligations* and *permissions*, those formal institutions can be neatly reflected in the widely applied tripartite structure of deontic logic [5].

In the context of computation, this static mapping has largely been adopted for the representation of informal insti-

tutions, such as social norms (e.g. [6]). In this work we pose a question that reaches beyond previous work, namely whether the static mapping of obligations and prohibitions offers a smooth alignment with informal interaction that mediates the transmission of conventions and social norms. Since the abundance of human interaction occurs in the informal realm, such as instructing subordinates, nurturing children, or sharing behavioural expectations, this mapping may be less clear cut and carry contextual meaning. In addition to general social-psychological considerations, the interpretation of behavioural conventions further includes specific cultural components [7] that influence individuals’ behaviour to a varying extent.

If we want to develop a grounded understanding in how far words can be used to represent different levels of normative prescriptiveness, e.g. in the form of mild inclinations (‘You might want to submit this work tonight.’) or by urging an action (‘You should submit this work tonight.’), inquiries into a shared understanding need to consider that semantics are primarily inferred from language use as opposed to ex ante specification [8]. We therefore rely on empirical investigations to model and represent the varying degrees of prescriptions in a computationally accessible model.

In its simplest form, linguistically prescriptiveness is reflected in modals that are associated with directives [9], such as ‘should’ or ‘must’, which we use as a basis to characterise levels of prescriptiveness. In this work we outline a methodology that affords the characterisation of commonly used directives with respect to their prescriptiveness, and thus shapes a computational vocabulary to characterise social norms accurately.

In the following Section II we outline previous work into the human understanding of deontic concepts. In Section III we introduce the conceptual overview of a continuous norm concept along with a methodological approach (Section IV) that includes survey design, term selection and initial analysis. In Section V we further provide insight into preliminary results of the ongoing study. Finally, in Section VI we plot out current limitations and perspectives for future refinements.

II. HUMAN UNDERSTANDING OF PRESCRIPTIONS

Norm concepts conventionally rely on deontics to capture possible choice on the part of the target of behavioural prescriptions. The essential concepts include ‘FORBIDDEN’ – specification of forbidden actions –, ‘OBLIGATORY’ – specification of mandatory actions –, and ‘PERMITTED’ –

¹Although these categories of institutions often overlap, this is by no means necessity, such as laws without enforcement and/or normative backing, which Ostrom [3] defines as *rules in form* (in contrast to *rules in use*).

specification of permitted actions. Those concepts provide a tripartite structure of interdefinable concepts [10]. For example, actions that are neither forbidden nor obligatory can be interpreted as allowed. In addition to those elementary deontic concepts exist the ‘NORMATIVE’ that captures the pre- and proscriptive deontics, i.e. all things that are obligatory or forbidden. On the other hand, the ‘DISCRETIONARY’ describes all things that are neither obligatory nor forbidden, thus emphasising action choices an individual has, beyond a mere permission. The relationships and semantics of deontics reach back into Aristotelian times and the complementarity of the deontic concepts has been reflected in the ‘deontic square of opposition’ (see e.g. [11]) as reflected in Figure 1.

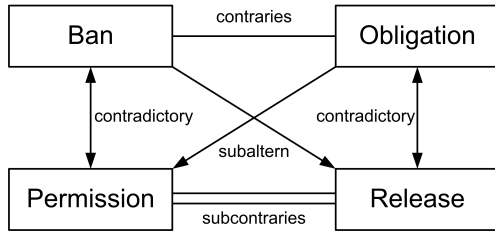


Fig. 1. Deontic Square of Opposition

The square captures the interrelationships and structures those into ‘contraries’ that stand in opposition to each other, such as prohibitions (‘Ban’) and obligations. Only one of those can apply at any given time. Contradictories capture terms pairs such as prohibition and permission or obligation and release, selected combinations of which, represented as subalterns, can be true at a given time. For example, an obligation relies on the permission to take an action, while a ban requires the release from an obligation. Similar for contraries, only one subcontrary can apply at any given time. These relationships provide the basis for the interdefinability of deontics and thus for deontic logic [5] and reasoning. In doing so, deontic logic emphasises notions of obligations, prohibitions and captures the intermediary rights (Permissions), but make no explicit reference to the subalterns (the actions of ‘releasing’ and ‘permitting’) that effectively represent prerequisites for the activation of and transition between normative states.

Earlier studies have explored the ability to perform deontic reasoning using the tripartite deontic concept and found high levels of accuracy [12], with particular precision with respect to the identification of rule violations [13]. Work by Cummins [14] further suggests that deontic reasoning is culture-independent and is neither sensitive to age groups.

The prototypical experimental paradigm for deontic reasoning tasks is Wason’s selection task [15], in which individuals are presented with rules and are asked to identify compliance or violation. While accuracy levels were high as long as explicit deontic clues were present [16], accuracy decreases with increasing abstraction of rules (e.g. symbolic representations) [17], and drops off to very low levels below 10 percent accuracy for tasks with weak deontic clues [18]. Beller [13] further suggests that the exact wording is secondary to individuals’ ability to infer its deontic meaning.

However, if we want to convey information of deontic value unambiguously, be it amongst humans, humans and computers, or computers, modals need to capture normative weight and

signal this with adequate precision to a message recipient (the *patient* in the context of linguistics). Norms are not exclusively characterised by distinct ‘musts’ and ‘must nots’, but likewise more moderate modals that en- or discourage behaviour, while maintaining the patient’s discretion over action choices.

In this work we thus intend to identify whether such refined deontic interpretations can be associated with specific terms, i.e. whether a deontic conception is attached to individual words. Moreover, this is associated with the question whether the deontic connotation of terms can be identified in a context-independent way and thus facilitate their generalisable use to express specific levels of normative prescriptiveness. As such, our aim is not to provide a comprehensive study, but to test whether a context-independent deontic understanding is reflected in terms used in everyday human language. To represent ‘oughtness’, we assume a continuous perspective on the deontics concept, which we explore in the following.

III. CONTINUOUS NORM CONCEPTIONS

To approach a fine-grained representation of prescriptiveness, we employ a continuous deontics conception that moves beyond a static identification of prohibitions, permissions and obligations, and instead pragmatically conceives those along a range with proscription and prescription at the extreme ends, the principle of which had already been considered by von Wright [19]. The concept introduced in this context has previously been introduced as a computational representation [20] and explored in simulations [21], [22]. However, in this context we borrow the concept to develop a grounded representation of human understanding that bears the potential to bridge the spheres of human and computational understanding of directives, such as in the context of suggestions, instructions or norms.

In this work we suggest that any directive can consequently be allocated along that range, permitting its association with the level of prescriptiveness. Bordering the permissive centre of this deontic range, the progression towards obligations would initially include the mere suggestion of particular actions and lead towards an increasing choice reduction on the part of the message recipient, such as the case for omissible activities, i.e. activities that can be foregone should circumstances require it, or if conflicting prescriptions of higher priority exist. Prescriptions ultimately culminate in obligations, leaving the recipient little space with respect to interpretation of directives and associated messages. This principle, both for prescriptive and proscriptive dimension is visualised in Figure 2 and tagged with terms² to exemplify the described progression from permissiveness towards pro-/prescriptiveness.

Besides the continuous nature of directives, the varying extent of prescriptiveness across the deontic range suggests a greater variability in terminological interpretation with possible impact based on cultural background or other demographic factors.

An assumption underlying this approach is the individualised mental representation of prescriptiveness that forms with exposure to social influence as well as further environmental

²Note that the choice of terms is not systematically grounded at this stage but are chosen to emphasize the progressive nature of allocated directives.

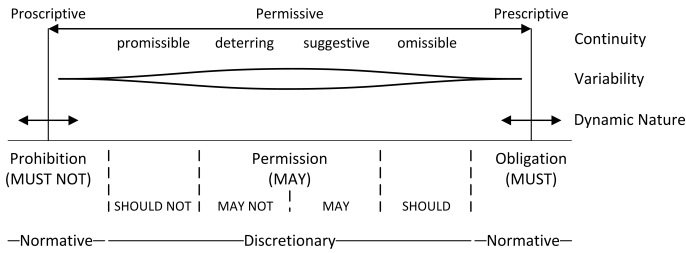


Fig. 2. Dynamic Deontics Conception

experience. The extent and breadth of influences inform the level of refinement that are allocated between *musts* and *must nots*, as ‘shades of grey’ between black and white. Examples of aspects that are influential in refining an individual’s normative understanding include the exposure to pluralistic societies as well as experiencing individuals from different cultures. A further influence on expansion or contraction of the deontic range is the extent to which those stimuli are sustained or subside, thus introducing a temporal component that affects the dynamics of the deontic range over time. Influences can further emphasise fine-grained representation of obligations or prohibitions over the other, potentially forming asymmetric deontic ranges. Consequently, the deontic ranges can vary between individuals, but are ultimately defined by the social influence sphere that shape an individual’s normative understanding. This approach permits the exploration on an individual level, while enabling the aggregation of deontic ranges on arbitrary levels of social organisation such as associating specific range distributions with societies or cultures.

In this work we take an initial step and attempt to map the human understanding of directives onto this scheme. In the following section we will provide a methodological outline before we proceed to preliminary results of this study.

IV. METHODOLOGY

Since our work yields towards an empirically grounded allocation of directives along the deontic range, we constructed a questionnaire that presents participants with a selection of directives that they classify from a perspective as message receiver based on three characteristics:

- *Pressure* – the perceived pressure exerted upon the message receiver (Question: “How much pressure would that term put on you as the receiver of the statement?”)
- *Variation* – the variation of this pressure depending on context, such as nature and status of message sender as well as situational setting (Question: “Do you think this pressure can vary in different cases or contexts?”)
- *Confidence* – the confidence a participant has with her/his choice (Question: “How confident are you with your choice?”)

The questionnaire employs continuous scales to capture participants’ choices. To accommodate a fine-grained evaluation of *Pressure* and *Variation* we use a comparatively fine-grained 11-point scale, while *Confidence* is indicated on a 7-point scale.

Participants are encouraged to leave comments related to the assessment of individual directives as well as the overall survey throughout the process.

In order to identify response patterns, the survey includes a set of demographic variables participants are encouraged (but not required) to provide. Those include Gender and Age. Particularly the latter aspect is treated with great care. Since the survey concentrates on the English language, survey participants are further asked to provide information about their mother tongue, country of origin, country of residence, and duration of residence in order to identify relationships between societal context and cultural exposure.

Furthermore, to preempt the confounding influence of respondent fatigue [23], the questionnaire exists in two randomly assigned variants, with original and inverted order of the core questions regarding the directive classification.

The sourcing of participants occurred via social networking sites as well contacting individuals directly.

A further aspect relates to the selection of directives used as part of the initial study. A challenge lies in the selection of terms that find use in the English language. The initial selection comprised terms used for the requirement specifications in IETF RFCs (as specified in [24]) and was used to inform the examples chosen in Figure 2. This selection has further been extended with related common terms (e.g. being expected to, have to), as well as synonyms and antonyms. From this body, a subset of terms was selected for the initial study, all of which are shown in Figure 3.

allowed	may not
appropriate	might
can	must
compulsory	must not
could	obliged
encouraged	permitted
essential	prohibited
expected	recommended
forbidden	refrain
have to	required to
important	should
inappropriate	should not
may	

Fig. 3. Selected Terms

Analytical aspects are described in the context of the initial analysis.

V. PRELIMINARY RESULTS

This initial survey was run with three objectives in mind. Literature (as well as the intuition of the authors) suggests the partial or complete context-dependence of word use as championed by Barsalou [25]. A primary concern is thus to identify 1) whether directive terms can be organised by prescriptiveness after all, and if so, to provide an outline of this progression. Further aspects involve the 2) refinement of the methodological approach and questionnaire as well as 3) to inform the term selection for future iterations.

A. Statistical Overview

The initial study received 66 complete responses, two thirds of which came from female participants. For this study the participation furthermore concentrated on Australian and New Zealand residents, with a distribution shown in Table I.

TABLE I. COUNTRY OF RESIDENCE

Country	Percentage
Australia	25.9 %
New Zealand	62.1 %
Other	12.0 %

The age distribution the participants ranges between 16 and 57 years, with a mean value of 28.86 years, median of 26 years, and a standard deviation of 9.3 years. Figure 4 illustrates the age distribution.

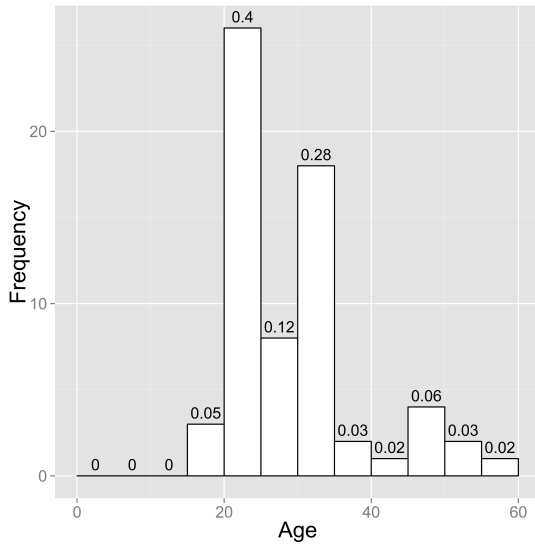


Fig. 4. Age Distribution

B. Term Evaluation

This leads us to discuss the core objective of this study, namely to identify whether directives can be structured with respect to their prescriptiveness, even if not explicitly specifying an exemplary context. Since the fine-grained 11-point scales are annotated with numeric values, we can apply statistical operations such as mean and standard deviation as shown in Table II.³ For nearly all terms we can identify a comparatively homogeneous spread, with standard deviations between 2 and 3 for the fine-grained 11-point scale and values around 1 for the 7-point confidence reading. However, we can observe greater spread with respect to the mean values for *Pressure* and *Variation*, while mean *Confidence* levels across all terms are comparatively constant.

To test the interaction between *Pressure* and *Variation* we used the Spearman correlation.⁴ The results (shown in

³Note the essential difference between Variation and standard deviation. Standard deviation expresses the diversity of value choices between different participants, while Variation emphasises the context-dependent value range ascribed to a particular term by individual participants.

⁴We rely on non-parametric tests since the collected data were not normally distributed.

TABLE II. DIRECTIVE TERM STATISTICS OVERVIEW

Term	Pressure		Variation		Confidence	
	mean	std. dev.	mean	std. dev.	mean	std. dev.
allowed	5.03	2.817	4.606	2.817	4.682	1.069
appropriate	6.015	2.49	4.667	2.507	4.364	1.047
can	3.47	2.309	5.091	2.664	4.53	0.996
compulsory	6.697	2.666	4.061	3.033	4.803	1.011
could	5.273	2.681	4.318	2.149	4.485	1.056
encouraged	4.303	2.795	4.515	2.457	4.561	1.04
essential	5.727	2.765	4.424	2.548	4.591	1.007
expected	6.303	2.155	4.682	2.121	4.561	0.979
forbidden	8.424	2.008	2.576	2.655	4.955	0.983
haveTo	7.621	1.975	4.864	2.728	4.667	0.791
important	6.091	2.154	5.576	2.512	4.621	0.799
inappropriate	6.091	2.21	4.667	2.394	4.227	1.25
may	2.5	2.099	4.667	2.926	4.606	1.051
mayNot	4.136	2.366	4.985	2.37	4.545	1.112
might	4	2.379	4.97	2.273	4.53	1.026
must	6.318	2.813	4.409	2.956	4.682	0.931
mustNot	7.167	1.886	4.303	2.643	4.591	1.052
obliged	7.773	1.983	3.833	3.076	4.682	1.025
permitted	4.333	2.477	4.333	2.544	4.561	1.01
prohibited	7.485	2.342	4.015	2.826	4.727	0.985
recommended	5.727	2.415	5.439	2.62	4.773	0.941
refrain	6.288	2.332	4.5	2.813	4.515	1.07
requiredTo	4.924	3.065	5.606	2.919	4.682	1.01
should	7.5	1.994	3.833	2.721	4.727	0.937
shouldNot	6.455	1.947	5.197	2.476	4.606	0.943

Table III) reveal negligible negative correlation levels between *Pressure* and *Variation*, and *Variation* and *Confidence*, respectively.⁵

TABLE III. CORRELATIONS BETWEEN PRESSURE, VARIATION, AND CONFIDENCE

Measure 1	Measure 2	Correlation	p-value
Pressure	Variation	-0.192	3.074×10^{-15}
Variation	Confidence	-0.168	5.230×10^{-12}
Pressure	Confidence	0.144	3.454×10^{-9}

Mapping the terms on the deontic range conception by associating disinclining terms with the prohibitive side of the deontic range and suggestive terms with the obligatory side respectively, provides the basis for a more comprehensive insight into the relevance of particular terms. Figure 5 shows the relationship between *Pressure* and *Variation* for mean values of directive terms plotted across the deontic range.

The plot reveals a bipolar orientation of terms, an aspect that can in part be associated with the poor exploitation of the extremal values of the 11-point scale by participants. A further suggestion that points towards extensions for future research could be a cultural component such as uncertainty avoidance [27] that drives decisiveness and avoids indifference (i.e. low pressure to take behavioural decisions). However, at this stage we cannot support such claims since our data base does not capture a sufficient amount of cross-cultural information.

Another noteworthy observation is the outlier ‘Forbidden’. In contrast to all other terms, ‘Forbidden’ appears to be surprisingly unambiguous with very low variation, while all other terms are associated with stronger context-dependence and thus higher variation.

The relationship between *Variation* and *Confidence* shown in Figure 6 supports this interpretation. Ascribing directive

⁵Our interpretation of correlation values follows [26].

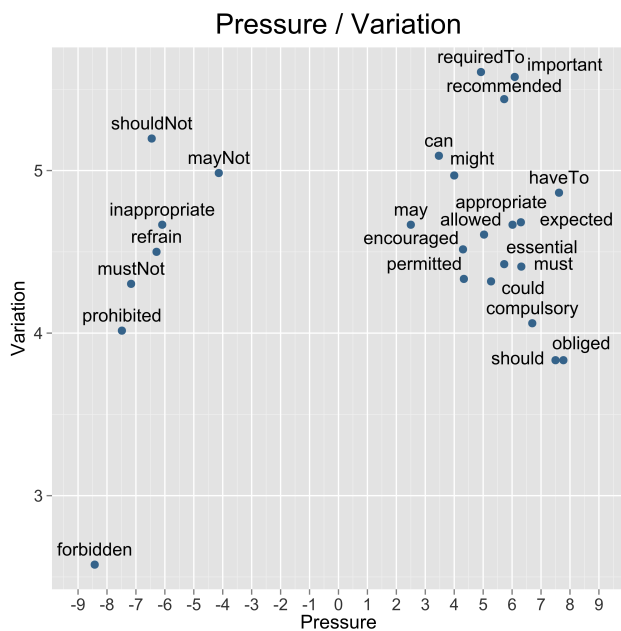


Fig. 5. Relationship between Pressure and Variation

terms greater variability offers a weak but noticeable correlation with reduced confidence.

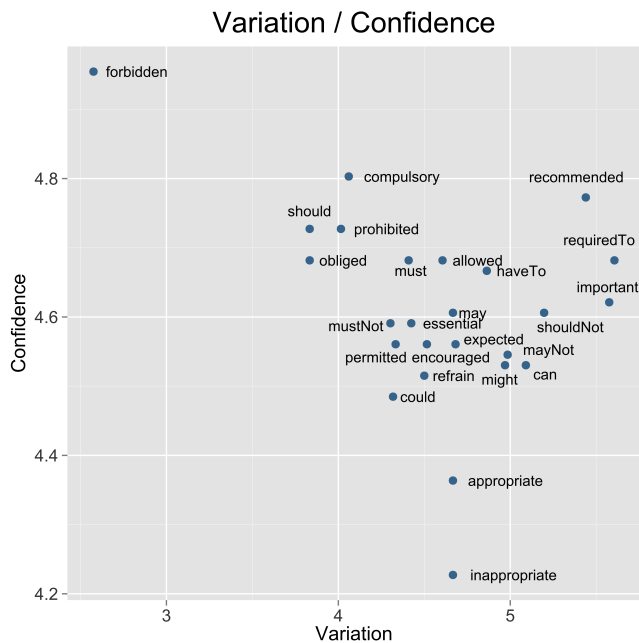


Fig. 6. Relationship between Variation and Confidence

The relationship between *Pressure* and *Confidence* as shown in Figure 7 complements to the relationship between *Pressure* and *Variation* (see Figure 5): Increasing prescriptiveness correlates with confidence; participants have an increasingly crisp understanding of what level of prescriptiveness the term entails.

However, independent of the strong levels of variation and limited direct correlation between *Pressure* and *Variation*, the

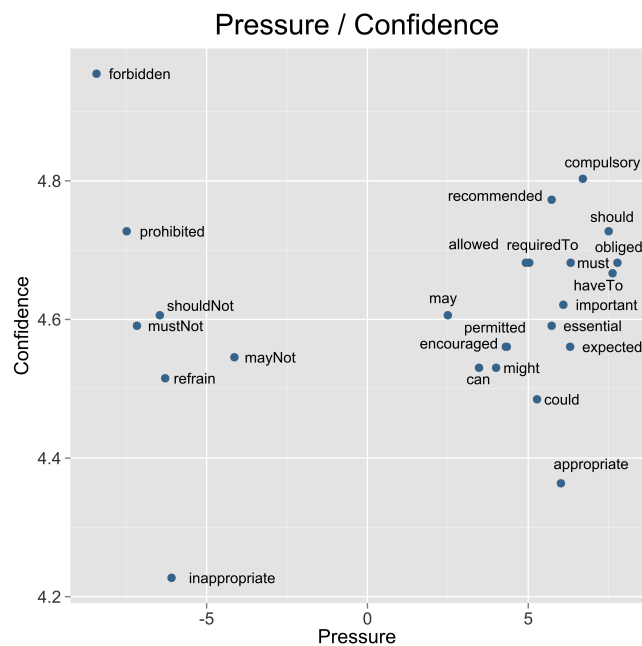


Fig. 7. Relationship between Pressure and Confidence

results show that the context-independent association of terms with different levels of pressure can at least be suggested. Note that our interpretation of context-independence is based on the assumption that the context-dependent interpretation varies between individuals, and therefore ultimately leading to a generalisable interpretation that does not rely on any specific context, since individuals are assumed to apply individualised exemplary contexts. Using this approach the study attempts to isolate the respective context-independent component (see Barsalou [25]) of directive terms.

Analysing the directive term progression used in our original visualisation in Figure 2, ranging from ‘must not’ towards ‘should not’, ‘may not’, ‘may’, ‘should’ and ‘must’, we can see this ordering loosely reflected in the data (see Figure 5). The only deviation from that order is the inverted ordering of ‘should’ and ‘must’, indicating a stronger level of obligation associated with the former. Moreover, the ordering of terms along the pressure dimension highlights a drop-off in variation towards either end of the deontic range; the stronger the prescriptiveness, the lower the variation, suggesting a less unambiguous interpretation. As such the results provide an indicative ordering of the considered terms with respect to their prescriptiveness.

However, the alternative dimensions variability and confidence offer the basis for a refined interpretation of terms with respect to their ability to accurately represent a desired level of prescriptiveness. Examples for this are the terms ‘have to’ and ‘obliged’. Though sharing similar levels of prescriptiveness (see Table II and Figure 5), the term ‘have to’ underlies stronger situational interpretation compared to ‘obliged’. In addition the multi-dimensional evaluation allows the classification of terms as equivalent with respect to their prescriptiveness. Examples include the densely clustered terms ‘appropriate’ and ‘expected’, as well as ‘should’ and ‘obliged’. Assuming the contextual applicability, those terms could be

used interchangeably to convey the same level of prescriptiveness.

Moreover, a multi-dimensional evaluation of terms by *Pressure*, *Variation*, and *Confidence* offers a basis for a refined comparison of directive terms, both to determine equivalence as well as varying levels of ambiguity (e.g. high context-dependence). We use this as a basis to inform the refinement of term selection for future studies.

C. Refining the Directive Term Selection

An essential feedback of this first study was that the number of terms participants were asked to characterise was considered too extensive. For further iterations of that study we thus intend to reduce the number of terms, but also to identify terms whose equivalence with other terms provides little added value and which can thus be treated as equivalent to evaluated terms and omitted from future evaluations.

To identify terms that can be substituted for future studies, we develop a systematic approach that allows us to discriminate given terms based on the dimensions *Pressure*, *Variation*, and *Confidence*. For all three dimensions we perform the Mann-Whitney-Wilcoxon test [28] for all term combinations with a significance level of 0.05. Since the essential interest lies in the determination of prescriptiveness, which is primarily associated with *Pressure*, we can exclude terms that offer sufficiently different *Pressure* levels from further analysis.

To isolate terms that offer similar levels of pressure, we introduce a staged identification of terms based on variation and confidence. As such terms of similar pressure levels could be differentiated based on their level of context-dependence and thus ambiguity, which in our case is represented as the contextual variation in pressure ascribed by participants. Terms with low levels of *Variation* thus indicate directives that can find generalisable context-independent use, in contrast to directive terms associated with high *Variation* levels, thus suggesting their strong context-dependence.

If not differentiated based on perceived pressure and context-dependence, a final potential means of discriminating terms is the *Confidence*, which is associated with the accuracy of classification by participants. Lower levels of confidence imply uncertainty about the value choice for *Pressure* and *Variation*, thus challenging the validity of the term classification. Terms with otherwise comparable measures but weaker confidence are candidates for substitution in future investigations.

Figure 8 illustrates the process of differentiating terms with respect to their deontic relevance.

Most terms explored in this work show significantly differing levels of prescriptiveness. In the following we thus concentrate on terms that bear similar pressure levels (α -level: 0.05), but either occupy the opposite sides of the deontic spectrum, vary in context-dependence, or are less authoritative based on participant confidence levels.

1) *Identifying Deontic Antonyms*: To identify terms that occupy the equivalent positions along the opposing end of the deontic range, we analyse terms that highlight an opposite valence, while bearing similar pressure and variation levels. Choosing terms with p-values > 0.5 offered the most refined results, all of which are shown in Table IV.

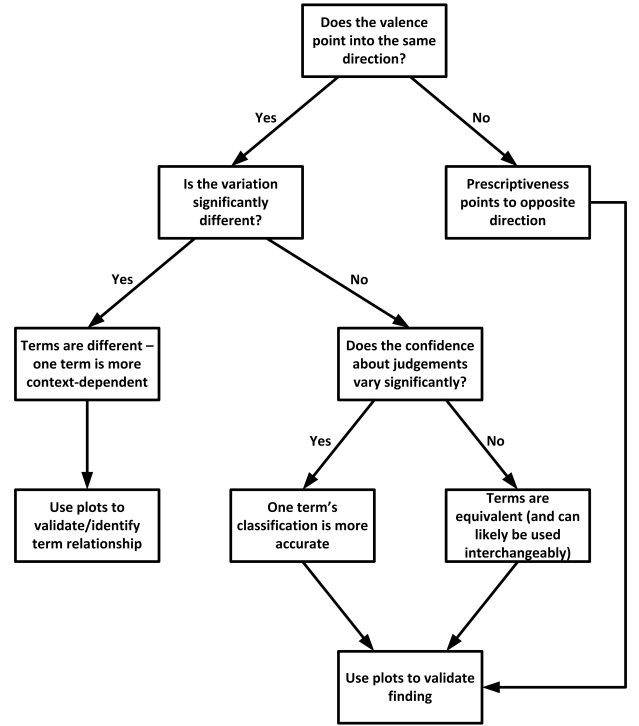


Fig. 8. Differentiating Terms with Insignificant Differences in *Pressure*

TABLE IV. SUGGESTED DEONTIC ANTONYMS

Directive Terms		Pressure	p-values Variation	Confidence
Term 1	Term 2			
obliged	prohibited	0.571	0.707	0.789
appropriate	inappropriate	0.978	0.987	0.6
appropriate	refrain	0.569	0.592	0.311
inappropriate	expected	0.551	0.971	0.163
inappropriate	essential	0.575	0.515	0.098
might	may/Not	0.659	0.947	0.745
should	prohibited	0.785	0.719	0.9
expected	refrain	0.894	0.512	1
refrain	must	0.681	0.9	0.402

The results bear little surprise. ‘obliged’ and ‘prohibited’ exist in distinct opposition, which is in line with previous studies that focus on extremal deontics [12]. The study can further identify ‘inappropriate’ and ‘appropriate’ as antonyms. The match of ‘should’ and ‘prohibited’, on the other hand, indicates that the prescriptiveness difference between ‘should’ and ‘obliged’ is minimal at best. A similar case holds for the term ‘refrain’ that is identified as antonym of ‘appropriate’. We will turn the identification of those ambiguous terms at a later stage.

Another aspect beyond the isolation of antonyms is the consideration of context-dependence. As indicated above, differences in variation highlight to what extent the prescriptiveness of a term depends on the context.

2) *Differentiating Terms by Context-Dependence*: To isolate terms of similar prescriptiveness with respect to their context-independent use, we extract the combinations of all terms that feature insignificant differences in pressure but significant differentiation in variation. Since variation indicates context-dependent use of terms, terms with significantly lower variation can thus be classified as less context-dependent, and make those terms candidates for a more general use. Extracted

terms of similar pressure, yet varying context-dependence are shown in Table V, with the column ‘General Term’ indicating the more context-independent of the compared terms.

TABLE V. SIMILAR TERMS OF VARYING CONTEXT-DEPENDENCE

Directive Terms		Pressure	p-values Variation	Confidence	General Term
Term 1	Term 2				
haveTo	obliged	0.514	0.034	0.818	2
haveTo	should	0.687	0.025	0.691	2
requiredTo	encouraged	0.272	0.014	0.505	2
requiredTo	allowed	0.793	0.038	0.983	2
requiredTo	essential	0.112	0.012	0.552	2
requiredTo	permitted	0.32	0.006	0.458	2
requiredTo	could	0.467	0.004	0.277	2
important	appropriate	0.987	0.036	0.165	2
important	compulsory	0.104	0.003	0.188	2
important	expected	0.509	0.016	0.771	2
important	must	0.307	0.016	0.458	2
important	essential	0.586	0.008	0.998	2
important	could	0.077	0.002	0.613	2
must	recommended	0.106	0.047	0.602	1
essential	recommended	0.863	0.031	0.294	1
recommended	could	0.334	0.014	0.122	2

As observed from Figure 5, prescriptive terms of similar pressure are largely organised in two clusters of similar pressure levels, with ‘required to’, ‘important’, and ‘recommended’ being most distinctively separated from the larger central cluster (including terms such as ‘could’, ‘allowed’). Terms such as ‘have to’, ‘obliged’, and ‘should’ show varying context-dependence, but are more prescriptive than the clusters mentioned before. Less prescriptive terms such as ‘can’ and ‘may’ are sufficiently differentiated with respect to their prescriptiveness and variation to prevent possible substitution despite varying levels of context-dependence.

D. Discriminating Terms of Similar Prescriptiveness and Context-Independence by Authoritativeness

Since we have clustered similar terms with respect to their context-dependence, terms can further showcase similar levels of prescriptiveness, while, at the same time, also exhibit similar levels of context-independence. To that stage those terms could thus be used interchangeably.

Since participants indicated levels of confidence with their choice, we can apply those to discriminate otherwise similar terms with respect to their authoritativeness. For such terms we could thus determine a clear preference. Table VI shows terms that are comparable with respect to prescriptiveness and context-dependence, but vary significantly with respect to participant confidence.

TABLE VI. IDENTIFIED TERMS OF GREATER AUTHORITATIVENESS

Directive Terms		Pressure	p-values Variation	Confidence	Preferable Term
Term 1	Term 2				
appropriate	compulsory	0.116	0.21	0.019	2
appropriate	recommended	0.42	0.11	0.021	2

In our context we can only identify ‘appropriate’ as a term whose characterisation was more challenging for participants than for the alternative terms ‘compulsory’ and ‘recommended’ as observable from Figure 6. Though classified with even lower confidence, the term ‘inappropriate’ is sufficiently different with respect to prescriptiveness and context-dependence to prevent substitution along the prescriptive part of the deontic range.

A final challenge of this analysis is to identify terms that are equivalent with respect to specific dimensions.

E. Identifying Ambiguous Terms

Given our relatively small sample and large standard deviations for evaluated terms, equivalence tests such as Yuen’s T test [29] did not resolve statistically significant equivalent terms. Though not classified as equivalent, in Table VII we nevertheless highlight the terms of closest relationship (p-values < 0.6 for *Pressure* and *Variation*), with lower p-values pointing towards greater similarity with respect to the respective dimension.

TABLE VII. EQUIVALENCE VALUES FOR SELECTED TERMS

Directive Terms		Pressure	p-values Variation	Confidence
Term 1	Term 2			
important	recommended	0.541	0.569	0.427
encouraged	permitted	0.39	0.538	0.018
allowed	could	0.401	0.453	0.382
appropriate	expected	0.379	0.274	0.176
appropriate	must	0.551	0.526	0.61
appropriate	essential	0.532	0.569	0.284
inappropriate	refrain	0.394	0.599	0.374
compulsory	must	0.458	0.552	0.247
expected	must	0.413	0.547	0.174

The terms with strongest approximation of equivalence are ‘appropriate’ and ‘expected’. Furthermore, the terms ‘allowed’ and ‘could’ showcase comparatively low p-values for both *Pressure* and *Variation*. Most other values show stronger alignment with respect to a particular dimension (*Pressure* for ‘encouraged’/‘permitted’ and ‘expected’/‘must’), but do not support any claim of equivalence. This suggests that the terms chosen for this exploration provide a sufficiently strong differentiation to avoid ambiguity. However, an extended exploration should seek further support for this finding and consider further terms beyond the selection presented in this work.

VI. SUMMARY, DISCUSSION, AND OUTLOOK

This work has attempted to address a distinctive question: Can we develop a generalisable differentiation of directive terms and organise those with respect to their prescriptiveness? The essential underlying hypothesis is that each directive term bears a sufficient context-independent component that permits its generic application. The underlying continuous deontics conception then facilitates the situation-specific mapping of terms to express the desired level of prescriptiveness, facilitating the computational representation of suggestions, encouragement, or mild disinclinations instead of falling back to a rigid tripartite norm conception that is defined in terms of obligations, prohibitions and a fuzzy permission concept.

The initial study introduced here proposes a methodology that allows us to develop an empirically grounded understanding of directives with respect to their prescriptiveness. At this stage the study shows a limited sample size and is thus inconclusive with respect to concrete results. However, the methodology facilitates the differentiation of terms by context-dependence and authoritativeness based on empirical grounding. In conjunction with the potential identification of antonyms, the analysis process bears intrinsic properties that inform a systematic refinement of term selection – not to

remove terms from selection, but to minimise the questionnaire lengths while maintaining the ability to systematically relate selected terms of individual study iterations for greatest possible coverage of the linguistic and deontic spectrum.

The shortcomings of the current study include a biased selection of terms; most terms capture the prescriptive spectrum of the deontic range, with few terms reflecting the proscriptive counterpart. Furthermore, the centre of the deontic range, capturing permissiveness, and generally associated with the term ‘may’, is hardly utilised for the given terms. Further studies should explore this phenomenon, both by extending the possible selection of terms to be classified as well as performing the study with larger numbers of participants to develop a more representative picture.

An important future endeavour is the establishment of measuring inter-cultural differences of the allocation of directives along deontic ranges. However, given our limited initial sample the analysis of cultural influences is premature, since the sample bears a strong bias towards New Zealand as country of residence, followed by Australian participants. Furthermore, both countries represented in the study share similar cultural characteristics as expressed in Hofstede’s Cultural Dimensions [27]. Sampling from a wide range of cultural backgrounds promises not only the basis to differentiate terms based on specific backgrounds, but allows us to identify the impact of specific cultural components, such as the relationship of collectiveness to norm understanding. Expanding the study furthermore offers grounds to identify in how far this approach measures normative understanding, in contrast to English language competence.

Notwithstanding these shortcomings for this iteration, our initial explorations support the hypothesis that a context-independent characterisation of directives with respect to their prescriptiveness is in principle possible. Application fields of a more refined study extend from the mediation of inter-cultural communication, such as choosing the right terms for the right target group, to human-computer interaction by providing machines with a refined understanding of prescriptiveness – enabling smooth interactions by avoiding terms that depend on social context but recommending terms that appropriately reflect the situational levels of desired expressiveness.

REFERENCES

- [1] E. Ullmann-Margalit, *The Emergence of Norms*, ser. Clarendon Library of Logic and Philosophy. Oxford (UK): Clarendon Press, 1977.
- [2] D. C. North, “Institutions,” *Journal of Economic Perspectives*, vol. 5, no. 1, pp. 97–112, 1991.
- [3] E. Ostrom, *Understanding Institutional Diversity*. Princeton (NJ): Princeton University Press, 2005.
- [4] W. N. Hohfeld, “Some Fundamental Legal Conceptions as Applied in Judicial Reasoning,” *Yale Law Journal*, vol. 23, pp. 16–59, 1913.
- [5] G. H. von Wright, “Deontic logic,” *Mind*, vol. 60, no. 237, pp. 1–15, 1951.
- [6] M. P. Singh, M. Arrott, T. Balke, A. K. Chopra, R. Christiaanse, S. Cranefield, F. Dignum, D. Eynard, E. Farcas, N. Fornara, F. Gandon, G. Governatori, H. K. Dam, J. Hulstijn, I. Krueger, H.-P. Lam, M. Meisinger, P. Noriega, B. T. R. Savarimuthu, K. Tadanki, H. Verhagen, and S. Villata, “The Uses of Norms,” in *Normative Multi-Agent Systems*, ser. Dagstuhl Follow-Ups, G. Andrighetto, G. Governatori, P. Noriega, and L. W. N. van der Torre, Eds. Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2013, vol. 4, pp. 191–229. [Online]. Available: <http://drops.dagstuhl.de/opus/volltexte/2013/4004>, accessed on: 1st June 2015
- [7] J. Henrich, “Does culture matter in economic behaviour? Ultimatum game bargaining among the Machiguenga of the Peruvian Amazon,” *American Economic Review*, vol. 90, no. 4, pp. 973–979, 2000.
- [8] L. Bloomfield, *Language*. New York (NY): Holt, 1933.
- [9] J. R. Searle, “A Classification of Illocutionary Acts,” *Language in Society*, vol. 5, no. 1, pp. 1–23, April 1976.
- [10] G. H. von Wright, *Norm and Action: A Logical Enquiry*. London: Routledge & Kegan Paul, 1963.
- [11] A. R. Anderson, *The logic of decision and action*. Pittsburgh, PA: University of Pittsburgh Press, 1956, ch. The formal analysis of normative systems.
- [12] S. Beller, “Deontic norms, deontic reasoning, and deontic conditionals,” *Thinking & Reasoning*, vol. 14, no. 4, pp. 305–341, 2008.
- [13] —, “A model theory of deontic reasoning about social norms,” in *Proceedings of the Twenty-Third Annual Conference of the Cognitive Science Society*, J. D. Moore and K. Stenning, Eds. Mahwah, NJ: Lawrence Erlbaum, 2001, pp. 63–68.
- [14] D. D. Cummins, “Evidence for the innateness of deontic reasoning,” *Journal of Memory and Cognition*, vol. 24, pp. 823–829, 1996.
- [15] P. C. Wason, *New Horizons in Psychology*. Harmondsworth: Penguin, 1966, ch. Reasoning, pp. 135–151.
- [16] R. A. Griggs and J. R. Cox, “The effects of problem content and negation on Watson’s selection task,” *Quarterly Journal of Experimental Psychology*, vol. 35A, pp. 519–533, 1983.
- [17] P. W. Cheng and K. J. Holyoak, “On the natural selection of reasoning theories,” *Cognition*, vol. 33, pp. 285–313, 1989.
- [18] P. C. Wason, “Reasoning about a rule,” *Quarterly Journal of Experimental Psychology*, vol. 20, pp. 273–281, 1968.
- [19] G. H. von Wright, *An Essay in Modal Logic*. Amsterdam: North Holland, 1951.
- [20] C. Frantz, M. K. Purvis, M. Nowostawski, and B. T. R. Savarimuthu, “Modelling Institutions using Dynamic Deontics,” in *Coordination, Organizations, Institutions and Norms in Agent Systems IX*, ser. Lecture Notes in Artificial Intelligence, T. Balke, F. Dignum, M. B. van Riemsdijk, and A. K. Chopra, Eds., vol. 8386. Berlin: Springer, 2014, pp. 211–233.
- [21] C. Frantz, M. K. Purvis, B. T. R. Savarimuthu, and M. Nowostawski, “Modelling the Impact of Role Specialisation on Cooperative Behaviour in Historic Trader Scenarios,” in *Coordination, Organizations, Institutions and Norms in Agent Systems X*, ser. Lecture Notes in Artificial Intelligence, N. Oren and A. Ghose, Eds. Berlin: Springer, 2015, in press.
- [22] —, “Modelling Dynamic Normative Understanding in Agent Societies,” in *Principles and Practice of Multi-Agent Systems - 17th International Conference, PRIMA 2014*, ser. Lecture Notes in Artificial Intelligence, H. K. Dam, J. Pitt, Y. Xu, G. Governatori, and T. Ito, Eds., vol. 8861. Berlin: Springer, 2014, pp. 294–310.
- [23] P. Ben-Nun, *Respondent Fatigue*, 0th ed. Sage Publications, Inc., 2008, pp. 743–744. [Online]. Available: <http://dx.doi.org/10.4135/9781412963947>, accessed on: 1st June 2015
- [24] S. Bradner, “Key words for use in RFCs to Indicate Requirement Levels,” <https://www.ietf.org/rfc/rfc2119.txt>, 1997, accessed on: 1st June 2015.
- [25] L. W. Barsalou, “Context-independent and context-dependent information in concepts,” *Memory & Cognition*, vol. 10, no. 1, pp. 82–93, 1982.
- [26] D. E. Hinkle, W. Wiersma, and S. G. Jurs, *Applied Statistics for the Behavioral Sciences*, 15th ed. Boston (MA): Houghton Mifflin, 2003.
- [27] G. Hofstede, G. J. Hofstede, and M. Minkov, *Cultures and Organizations: Software of the Mind*, 3rd ed. New York (NY): McGraw-Hill, 2010.
- [28] H. B. Mann and D. R. Whitney, “On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other,” *Annals of Mathematical Statistics*, vol. 18, no. 1, pp. 50–60, 1947.
- [29] K. K. Yuen, “The two-sample trimmed t for unequal population variances,” *Biometrika*, vol. 61, no. 1, pp. 165–170, 1974.