“Must” people reason logically with “permission” in daily situations? An explorative experimental investigation in human reasoning of normative concepts.

Wai Wong (wai.wong@kuleuven.be)
Meimei Yang (yangmeimei.gdufs@gmail.com)
Walter Schaeken (walter.schaeken@kuleuven.be)
Laboratory for Experimental Psychology, KU Leuven

Lorenz Demey (lorenz.demey@kuleuven.be)
Centre for Logic and Philosophy of Science, KU Leuven

Joost Vennekens (joost.vennekens@kuleuven.be)
Department of Computer Science, KU Leuven

Abstract

Philosophers have long been arguing the precise semantics of different deontic terms within normative statements. However, little research has been done on the human reasoning side of understanding such terms. In this paper, we propose a normative scheme with bitstring semantics that is expressive enough to cover the basic normative concepts in most mainstream schemes proposed in deontic logic research. Even though further confirmation is needed, our explorative experiments on human deontic reasoning have shown results that are consistent with our proposed scheme.

Keywords: bitstring semantics; deontic logic; deontic reasoning; experimental pragmatics; knowledge representation; normative concepts; scalar implicatures

Introduction

Normative statements are everywhere in the world and we encounter them every day, e.g. in rules and laws: *(A) Students may use laptops during the lecture; (B) During winter, there is an obligation for the caretaker that the room is warm enough; (C) In the trolley problem, one ought to change the track to kill one life in order to save five lives (Foot, 1967); (D) There is an obligation for the driver that their passenger wears a seatbelt; (E) There is an obligation for the borrower that the borrowed book is returned; (F) The war ought to be stopped.* The precise meaning of such statements has already been the subject of much debate in the literature. In this paper, we will first introduce basic normative concepts before we introduce our proposed scheme on deontic logic. Explorative experiments were conducted to see if the scheme can predict and explain our understanding of normative concepts.

Preliminary: Basic Normative Concepts

Normative statements contain many different components, with different origins and they involve different stakeholders. To easier understand our proposed scheme, we will first go through different basic normative concepts one by one. While we introduce several more concepts, the main focus of our paper will be on (co-)scope and deontic terms.

Scope In every normative statement, there is a *scope* to which a certain *normative sentiment* — be it praise, blame, or anything else — is attached. The *scope* can contain an action[1] (e.g. using the laptop in (A)) which involves an *actor* (e.g. student in (A)) or simply a state of the world[2] (e.g. the room is warm enough in (B)). Whenever what is stated in the *scope* happens, the attached *sentiment* will follow.

Co-scope Not only the presence of what is stated in the *scope* will bring out a *sentiment*, but also the absence of it. To quote Chisholm (1982): “*in considering how to classify the status of a given action, we must consider both the status of its performance and the status of its non-performance*” (McNamara, 2021). This contrast is also supported by the fact that human explanations are contrastive (Miller, 2019), Rather than focusing only on the *scope* prescribed by the normative statement, we will compare the *sentiment* of the *scope* with that of another scope. Here we coin a new term, *co-scope*, which is a contrasting scope that also has a certain *sentiment* attached to it. Usually the *co-scope* will be defined as the absence of the *scope*’s content or the opposite of it. However, under certain circumstances, the *co-scope* will be defined differently. For example, in the trolley problem (C), the *co-scope* contains “killing five lives by staying put”.[3]

Deontic terms In a normative statement, there is a deontic term indicating what *sentiments* apply to follow the *scope* and *co-scope*. By considering what *sentiments* are entailed by the *scope* and *co-scope*, we can classify deontic terms into different normative concepts. This approach is supported by Meinong who “*employ two contrary terms in combination to define the target normative concepts*” (McNamara, 2021). In

[1] This can be seen as a reflection of the “ought-to-do” concept in deontic logic. (Meinong, 1894)

[2] Similarly, this can be seen as a reflection of the “ought-to-be” concept in deontic logic. (Marcus, 1966)

[3] A similar concept can be seen in dyadic deontic logic (Parent & van der Torre, 2018) which introduces conditional in the *scope*. If conditional is defined as either *scope* or *co-scope*, then the *co-scope* would always be the opposite of the *scope* — “killing one life by changing track” in (C). Using the trolley’s problem (C) as an example, the conditional would be “either kill five lives by staying put or kill one life by changing track”.

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

the next section, we will propose a scheme that is expressive enough to represent common normative concepts and illustrate the relationship among *scope*, *co-scope* and *deontic terms* according to the proposed scheme.

Bearer The *bearer* is the one who is responsible for what happens, be it in the *scope* or *co-scope* of the normative statement, i.e. they are the one who is on the receiving end of the attached *sentiment*. Usually when the scope involves an action, the *bearer* and the *actor* will be the same agent (e.g. student in (A)). However, that is not necessarily the case as the *bearer* can also be different from the *actor* (e.g. driver as *bearer* and passenger as *actor* in (D)) or even there is no specified *actor* (e.g. borrower as *bearer* and anyone like borrower's friend can be as *actor* in (E)). Last but not least, *bearer* can also be unspecified as well (e.g. (F); see McNamara, 2004; Frijters, 2021; Frijters et al., 2021).

Normative source The *sentiment* that is attached to the normative statement usually comes from a specific *normative source*, such as one's moral judgement, rules within an organization, or the laws (Hage, 2017). There can be multiple *normative sources*, each having its own sets of normative statements and the resulting *sentiment* followed can be conflicting. The *normative source* can also be left unspecified. In this study, we assume there is only one *normative source* without any conflicts.

Static rules vs dynamic rules In the normative statement examples we have seen so far, we have only discussed rules (or normative statements) that are static. Moreover, there may also be rules that are about introducing, modifying or removing rules (or normative statements), especially in laws. Here are a few examples: *Book owners can decide who is allowed to read their book; Lecturers cannot remove the eating ban in the lecture hall; Teachers can make exceptions for the assignment obligation required of the students.* Rules that govern the introduction, modification or removal of rules are called dynamic rules whose *(co-)scope* contains also another normative statement. In contrast, static rule contains only a normal statement without normative flavour[4]. In this study, we limited our research to static rules.

Proposed Scheme for deontic terms

In this section, we are going to propose a new representation scheme for deontic terms. The main motivation behind this proposal is that the subtle differences between deontic terms often lead to conflation of meaning, and confusion on the normative statements. This conflation phenomenon — according to McNamara (2021) — hinders the development of logics of normative concepts. It is our hope that by introducing this scheme, we can express the subtle differences in the meaning of deontic terms well enough while still avoiding the confusion commonly seen when defining deontic terms.

Table 1: The 13 different atomic situations

Combination		A	B	C	D	E	F	G	H	I	J	K	L	M
Sentiment of *Scope*	+	×	×	×	×	○	○	×	×	○	○	○	×	×
	0	×	×	×	○	×	×	○	×	×	×	×	○	×
	-	○	○	○	×	×	×	×	○	×	×	×	×	○
Sentiment of *Co-scope*	+	×	×	○	○	○	○	×	×	○	×	×	×	×
	0	×	○	×	×	×	×	○	×	×	○	×	×	×
	-	○	×	×	×	×	×	×	○	×	×	○	○	○
Sentiment of *Scope* vs. *Sentiment* of *Co-scope*	<	○	○	○	○	○	×	×	×	×	×	×	×	×
	=	×	×	×	×	×	○	○	○	×	×	×	×	×
	>	×	×	×	×	×	×	×	×	○	○	○	○	○

Under this scheme, there are only three types of normative sentiment — praise (+), neutrality (0) and blame (-) and we assume praise is always more preferable than neutrality which is always more preferable than blame. We consider three basic questions:

What sentiment would follow if the scope is realised?
What sentiment would follow if the co-scope is realised?
Which sentiment of the scope and co-scope is preferred?

While there are three possible answers for each basic question, in total there are only 13 possible combinations of answers to the three basic questions due to the sentiment preference assumption (see Table 1). We view each of these 13 combinations as a unique atomic situation. A deontic term can then apply to a number of these atomic situations. We link deontic terms to bitstring[5] of length 13 — $\overline{ABCDEFGHIJKLM}$: a '1' in such a bitstring means that the deontic term applies to the corresponding situation, and vice versa for a '0'. If all '1's in a bitstring is observed in the same position in another bitstring, then former entails the latter.[6]

Among all normative concepts, supererogation, which involves surpassing the expected duty or doing more than required, is a well-known aspect of moral awareness. It belongs to a diverse group of related ideas that pose difficulties in being accurately represented together in deontic logic and ethical theory. In McNamara's (2021) overview of logical frameworks and schemes that address the concept of supererogation, we have identified five schemes, which can be expressed in our proposed scheme (see Table 2 and Figure 2). When representing the traditional threefold classification scheme, we make a conscious choice — on the one hand, in a situation where both *scope* and *co-scope* would bring blame (e.g. the trolley problem), one has a duty to be optimal, i.e. obligated to choose the one with a lesser degree of blame, and free to choose if both blame are of the same degree; On the other hand, in a situation where both *scope* and *co-scope* would bring praise, the duty to be optimal ceases to exist and one is free to choose. This interpretation in our opinion also aligns with human's risk-averse nature (Kahneman & Tversky, 1982) — every additional harm can significantly increase the threat to life, while opting for a lesser reward will not lead to the end of life. This

[4] In other words, static rules are defined as their (co-)scope which do not contain normative statements.

[5] This representation style follows the convention of the bitstring semantics framework, see Demey & Smessaert (2018).

[6] Other logical relationships such as contradictory (A bitstring has flipped all the '1's and '0's of another at the same position), contrary (shares no '1' at the same position) and subcontrary (shares no '0' at the same position) can be easily observed in bitstring semantics.

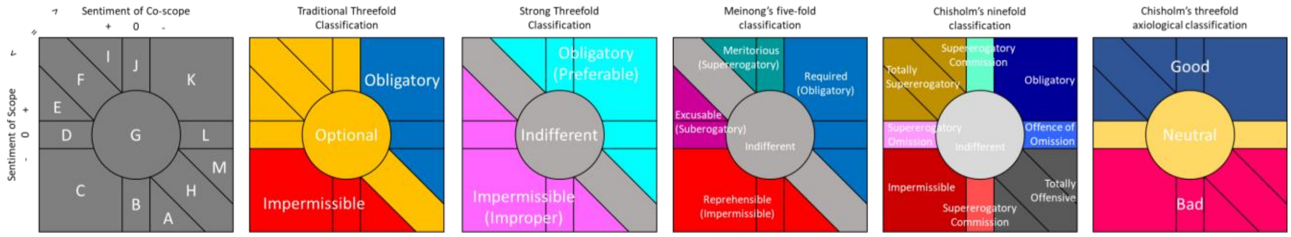

Figure 2: Graphical representation of different scheme

interpretation also leaves room for representing the concepts of supererogation (i.e. rather-so) and suberogation (i.e. better-not) in Meinong's finer five-fold classification, which extends the traditional threefold classification scheme by combining it with the strong threefold classification scheme.[7] Even if we have a well-defined framework for normative concepts, if humans do not reason as predicted in the framework, such a framework is of little use in explaining how humans reason deontic terms. As far as we know, the focus of the current psychological research on deontic reasoning (Beller, 2010) is developmental, i.e. how and when humans develop deontic reasoning skills rather than how humans understand different deontic terms. To bridge this gap, we conduct a new series of experiments.

Table 2: Bitstring semantics representation of different prominent deontic terms in major normative logical scheme

Scheme	Deontic term	Bitstring
Traditional Threefold Classification	Obligatory	0000000000111
	Permissible	0001111111111
	Impermissible	1110000000000
	Omissible	1111111111000
	Optional	0001111111000
	Non-optional	1110000000111
Strong Threefold Classification	Obligatory (Preferable)	0000000011111
	Impermissible (Improper)	1111100000000
	Indifferent	0000011100000
	Morally Significant	1111100011111
Meinong's five-fold classification	Meritorious (Supererogatory)	0001100000000
	Required (Obligatory)	1110000000000
	Indifferent	0000011100000
	Excusable (Suberogatory)	0000000011000
	Reprehensible (Impermissible)	0000000000111
Chisholm's ninefold classification	Totally Supererogatory	0000110010000
	Supererogatory Commission	0000000001000
	Obligatory	0000000000100
	Supererogatory Omission	0001000000000
	Indifferent	0000001000000
	Offence of Omission	0000000000010
	Impermissible	0010000000000
	Offence of Commission	0100000000000
	Totally Offensive	1000000100001
Chisholm's threefold axiological classification	Good	0000110011100
	Neutral	0001001000010
	Bad	1110000100001

Experiments

There are three rounds of experiments in our study. In the first round, we test the understanding of the six deontic terms in the traditional scheme. If humans reason according to our proposed scheme, there are four hypotheses: (1) people will agree with "obligatory" (disagree with "forbidden") when the content in the scope is realised, and vice versa; (2) on average, people will not disagree with "optional" no matter the situation; (3) people will also agree with "permitted" (disagree with "released") when the content in the scope is realised, and vice versa. In addition, given that both terms are entailed by "optional", people will tend to agree more with these terms than the stronger term "obligatory" ("forbidden"), and (4) people will only agree with "non-optional" either when the content in the scope is realised for every *bearer* (i.e. students in the scenario, see Figure 1) or when the content in the *co-scope* is realised for every *bearer*. In the second round, we will address the issue of meaning conflation in the strong scheme. In particular, we will look into the terms "preferable" and "uncalled for". If meaning conflation does exist (which implies only the strong scheme is valid, which goes against our proposed scheme as we also accommodate the traditional scheme), there is an additional hypothesis on the outcome of the experiment: (5) the trends we observed in "preferred" ("uncalled for") will be similar to that in "must" ("banned"). We also conducted a third round of experiments to test most of the hypotheses listed above once again with sentences that share a uniform structure. In total, we will investigated eight normative concepts (see Table 3). While the hypotheses we investigate here may not be new to philosophers, it will be the first time that those hypotheses are being put into test in the form of psychological experiment.

Table 3: List of normative concepts in the experiment.

Concept	Round 1	Round 2	Round 3
1110000000000	Forbidden	Banned	Prohibition
1111100000000	-	Uncalled for	Impropriety
1111111111000	Released	Exempted	Exemption
0001111111000	Optional	-	Option
0001111111111	Permitted	May	Permission
0000000011111	-	Preferable	Preference
0000000000111	Obligatory	Must	Obligation
1110000000111	Non-optional	-	-

[7] The strong scheme differentiates itself from the traditional scheme with the conflation of "optional" to "indifferent". To highlight such difference in our proposed scheme, we further assume that under the strong scheme one ought to be morally optimal, i.e. make a choice that brings bigger praise or lesser blame.

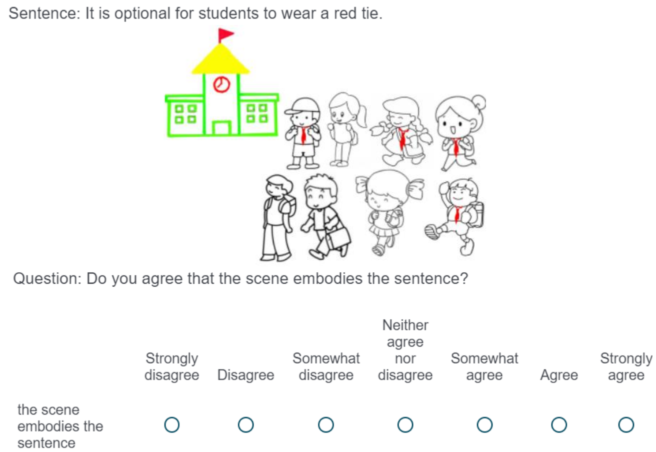

Figure 1: An example of the task format.

Participants

In each round of the experiment, 50 participants were recruited via the specialized online platform Prolific. There were 150 participants in total (mean age=31.27, range=18-45, 50% female). They were selected based on the following criteria: (a) are between 18 and 45 years old; (2) are native English speakers; (3) have experience in using computers, mobiles, or other similar electronic devices. Considering that the study aims to recruit native English speakers, we distribute the experiments in English-speaking countries.

Materials and Procedure

To test whether human reasoning follows our proposed logical scheme, we carry out sentence verification tasks which are commonly used in experimental pragmatics studies (e.g., Bott & Noveck, 2004; Noveck, 2001; Noveck & Posada, 2003). The trials in the experiment are presented in the form of a sentence containing a deontic term, a picture of a certain scenario and a question asking participants to indicate whether the sentence is a good description of the picture on a 7-point Likert's scale ranging from "strongly disagree" to "strongly agree". Figure 1 is an example of the task format. In each of the three rounds, a set of 6 or 7 sentences was being investigated:

Round 1 (conducted on 24/05/2023)

1. It is obligatory for students to wear a red tie.
2. Students are permitted to wear a red tie.
3. It is optional for students to wear a red tie.
4. Students are released from wearing a red tie.
5. It is forbidden for students to wear a red tie.
6. It is non-optional for students to wear a red tie.

Round 2 (conducted on 03/01/2024)

1. Students are exempted from wearing a red tie.
2. Students must wear a red tie.
3. Students may wear a red tie.
4. Students are banned from wearing a red tie.
5. Students wearing a red tie is preferable.
6. Students wearing a red tie is uncalled for.

Round 3 (conducted on 03/01/2024)

1. There is an exemption for students to wear a red tie.
2. There is an obligation for students to wear a red tie.
3. There is a permission for students to wear a red tie.
4. There is a prohibition for students to wear a red tie.
5. There is a preference for students to wear a red tie.
6. There is an impropriety for students to wear a red tie.
7. There is an option for students to wear a red tie.

The picture following the sentence showed 8 students in each scenario where either 0, 1, 3, 4, 5, 7 or 8 out of 8 students were wearing a red tie. In total, there were 7 scenarios and in each round of the experiment, each sentence was matched with a scenario, forming either 42 (round 1 & 2) or 49 (round 2) trials which appeared in a random order to the participants to avoid any bias caused by ordering. However, instead of fully randomizing all 42 items, we first put the 6 or 7 different statements in random order and then randomized the trials per statement. After these 42/49 trials, there was a self-evaluation question that participants were asked to evaluate their understandings of the tested deontic terms. In addition, the duration of each individual trial was also recorded.

Results and Discussion

In this experimental study, our main interest is the relationship between participants' levels of agreement towards the use of the various deontic terms and the number of students wearing a red tie in the shown scenarios. On the one hand, these relationships can be observed via bar plots (see Figure 3). On the other hand, these relationships can also be investigated via a mixed-effect regression model, which includes both linear and quadratic form of proportion (i.e., the number of students out of eight wearing a red tie) and their interaction with deontic terms as fixed-effect factor, to predict participants' levels of agreement. We also included a random-effect factor in the model to account for participants' general tendency to give a higher/lower score. We performed an ANOVA analysis on the significance of the fixed-effect factors. The data were sorted out and analyzed in R (R Core Team, 2023). Table 4 presents the test results that almost all the fixed-effect predictors are statistically significant. This model explains in total 61.7%, 56.3%, and 52.4% variance for the outcome, with 58.7%, 46.5% and 48.2% explained by the fixed effect and around 3%, 9.8% and 4.2% by the random effect of individual differences for round 1, 2 and 3 of experiment respectively. To visualize the model in a

Table 4: Significance test for the fixed-effect factors.

Fixed-effect factor	Pr(>F)		
	Round 1	Round 2	Round 3
Deontic Terms	<0.001	<0.001	<0.001
Proportion (Linear)	0.004	0.064	<0.001
Proportion (Quadratic)	<0.001	<0.001	0.044
Deontic Terms × Proportion (Linear)	<0.001	<0.001	<0.001
Deontic Terms × Proportion (Quadratic)	<0.001	<0.001	<0.001

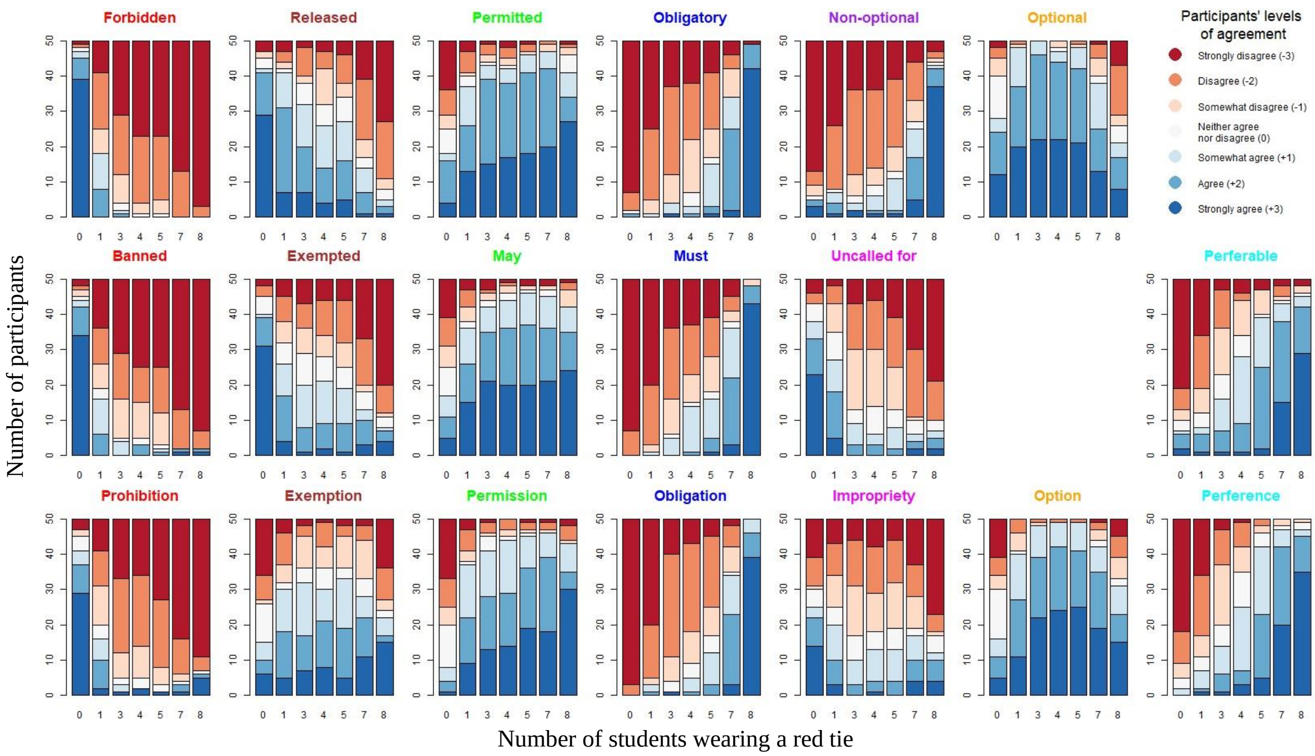

Figure 3: Bar plots of the number of participants in giving each score on different trials

straightforward way, we presented fitted lines in Figure 4 which simulates the relationship between the predictor of the proportion of students wearing a red tie and the expected outcome of participants' levels of agreement in considering the use of deontic term appropriate for the situation. Unsurprisingly, those fitted lines do reflect the relationships observed in the bar plots (Figure 3). While proportions of agreement levels may change within each term, participants also showed various levels of agreement towards the use of deontic terms no matter in the same or different scenarios.

As predicted by our hypotheses, in all three rounds of experiments with one exception, "obligatory" and "permitted" ("forbidden" and "released") display a trend that more and more (less and less) participants agreed with these terms as more red ties were worn, while the agreement level is higher in general for "permitted" ("released") than for "obligatory" ("forbidden"). In the exception case, for the concept of "omissible" ("released" / "exempted" / "exemption") in the round 3 experiment, rather than strictly decrease as in other rounds, agreement first increases and then decrease. This might be due to the sentence structure "there is…" implying that some but not necessarily all students have the "exemption".

A similar trend is seen for "optional", where the agreement first increased and then decreased, which is in line with our hypothesis. However, the trend in "non-optional" goes against what is predicted in the hypothesis and it closely follows the trend observed in "obligatory", which suggests a

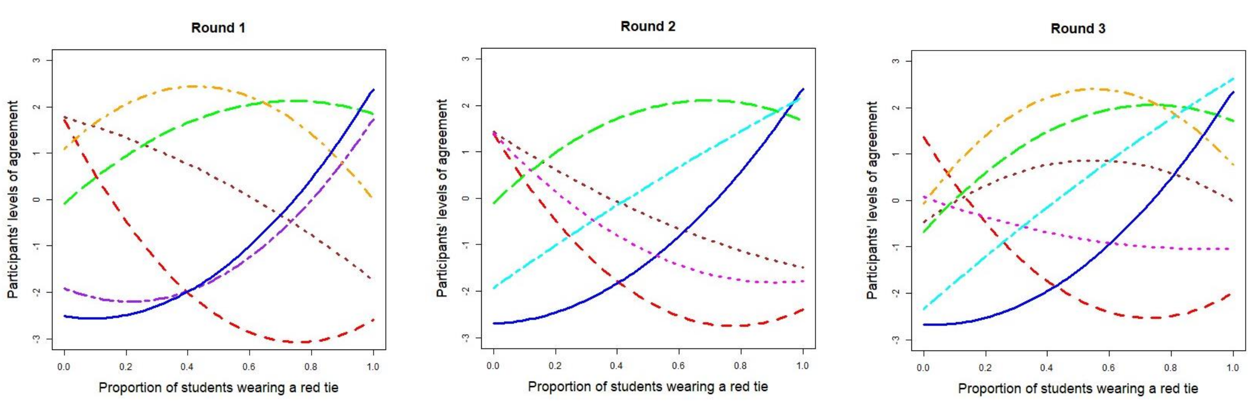

Figure 4: Fitted lines from regression models to predict participants' levels of agreement. The colour of each line indicates the corresponding deontic terms of which the same colour in table 3 (and in figure 3).

potential meaning conflation in action. Last but not least, meaning conflation is not observed for “impropriety” and “preference”, but rather the trend of “impropriety” is sandwiched between those of “forbidden” and “released” while the trend of “preference” is sandwiched between those of “obligatory” and “permitted”. This result is in agreement with the semantics of the deontic terms in our proposed scheme.

On the self-reported confidence question, nearly all participants reported a positive confidence level towards their understanding of all deontic terms with the exception of “released” and “impropriety”. Notice that the narrow range of agreement levels on “impropriety” (between 0 and -1) in round 3 might be due to the fact that participants found it difficult to understand. Nonetheless, most participants spent less than 20 seconds on each trial, regardless of the deontic terms.

Conclusion

So far our experimental results on human deontic reasoning are mostly consistent with our proposed scheme. However, human reasoning can sometimes deviate from what is expected in the scheme. In this section, we are going to discuss what are the possible causes of such deviation, what insights this explorative study has brought us for future research, and what is the implication of our proposed scheme in the development of deontic logic and knowledge representation of rules on human behaviours.

Meaning conflation and scalar implicatures

Meaning conflation is at the heart of argument and confusion in deontic logic research. As we see from the results, such conflation can be observed in some deontic terms, such as “non-optional”. One possible explanation for such conflation is a pragmatic phenomenon called scalar implicatures — the information expressed by the weaker term in a semantic scale (e.g., “non-optional”) is enhanced with the perceived rejection of the stronger term(s) belonging to that same scale (e.g., “forbidden”). As research has shown that such implicatures are associated with processing costs (Nys et al., 2024), in future research a dual-task approach (e.g. Janssens & Schaeken, 2016) can be applied to determine participants’ perceived literal meaning by overloading participants’ working memory so that no scalar implicatures are elicited.

Asymmetry in patterns of mirrored deontic terms

A mirrored deontic terms pair can be defined from each separate term by simply swapping the *scope* and *co-scope*, e.g. “Obligatory” & “Prohibited”, “Permitted” & “Exempted” and “Preference” & “Impropriety”. One would expect that the relationships shown in Figure 4 for each of those pairs would simply be a reflection of each other. However, there is asymmetry for all pairs, especially pronounced in the “Permitted” & “Exempted” pairs. However, this result is not surprising, as research (van Tiel et al., 2019) has shown that the processing cost is different for negative scalar words and positive scalar words. Also, the difference in the degree of asymmetry in different pairs can also be possibly explained by scalar diversity (Pankratz & van Tiel, 2021), i.e. the variation in the occurrence of scalar implicatures across different semantic scales.

Formalisation of deontic interpretation

Our proposed scheme has been shown to be expressive enough to represent all mainstream schema being proposed in the field of deontic logic. As we mentioned earlier, one challenge in deontic logic is meaning conflation. This issue is solved in our scheme via the bitstring annotations which clearly show which atomic situations are covered by the terms. However, the difference between deontic terms does not always lie in the bitstring semantics; sometimes it is in the normative sources. For example, we argue that while both “must” and “ought” can be represented as 0000000000111, “must” normally comes from sources like rules and laws; By contrast, “ought” comes from sources like “moral judgement” which usually hold a higher standard. By carefully identifying both the bitstring semantics and its corresponding normative source, it is possible to formalise the normative concepts and different deontic interpretations without ambiguity.

Knowledge representation of human rules

For computer system, *normative sentiment* is enough to guide them in executing the desired behaviour (e.g. reinforcement learning, see Kaelbling et al., 1996). However, most of the time a rule (or a normative statement) by itself is not enough to motivate people to follow it. As humans, we resort to attach real-world consequences to the *sentiment*, e.g. a reward following praise, a punishment following blame, etc., to incentivise people. In laws, there are clearly defined stakeholders who can enforce the rules by giving out real-world consequences (law enforcement, e.g. police officers, prison guards) and who can evaluate and determine which appropriate real-world consequences will follow (e.g. judges, juries). We speculate responsibilities of those stakeholders can still be well-defined by using just the normative concept components mentioned earlier in the introduction section. However, such research in logical formalizing of deontic responsibilities (which is not cover in this paper) remains to be seen as the research on deontic logic is underdeveloped (Meyer & Dignum, 1994; Benzmüller et al., 2018; Governatori et al., 2021), let alone the formalization of laws using deontic logic. To successfully represent the rules the governs human behaviours into a knowledge system, one cannot simply use the same normative system which governs computer behaviours that follow the rules to the letter; Rather they must address this underdeveloped research area and formalizing the rules that oversee the attachment of real-world consequence to the normative sentiment. As mentioned in the introduction, our proposed scheme can covers all the basic normative concepts with the potential to represent the dynamic rules as well. It is our hope that this research can act as a stepping stone towards the development in comprehensive knowledge representation of human rules.

Acknowledgments

The work of the corresponding author (Wai Wong) is currently supported by a doctoral grant from the Internal Funds KU Leuven (BITSHARE-project IDN/19/009). The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Beller, S. (2010). Deontic reasoning reviewed: psychological questions, empirical findings, and current theories. Cognitive Processing, 11(2), 123–132.

Benzmüller, C., Parent, X., & van der Torre, L. (2018). A deontic logic reasoning infrastructure. In *Sailing Routes in the World of Computation: 14th Conference on Computability in Europe, CiE 2018, Kiel, Germany, July 30–August 3, 2018, Proceedings 14* (pp. 60-69). Springer International Publishing.

Bott, L., & Noveck, I. A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language*, *51*(3), 437–457.

Chisholm, R. M. (1982). Supererogation and offence: A conceptual scheme for ethics. In Brentano and Meinong studies (pp. 98-113). Brill.

Demey, L., & Smessaert, H. (2018). Combinatorial bitstring semantics for arbitrary logical fragments. *Journal of Philosophical Logic*, *47*, 325-363.

Foot, P. (1967). The problem of abortion and the doctrine of the double effect.

Frijters, S. (2021). Quantification in Normative Reasoning. In *All doctors have an obligation to care for their patients : term-modal logics for ethical reasoning with quantified deontic statements* (pp. 23-46). Ghent University. Faculty of Arts and Philosophy.

Frijters, S., Meheus, J., & Putte, F. V. D. (2021). Reasoning with rules and rights: term-modal deontic logic. In *New Developments in Legal Reasoning and Logic: From Ancient Law to Modern Legal Systems* (pp. 321-352). Cham: Springer International Publishing.

Governatori, G., Rotolo, A., & Sartor, G. (2021). Logic and the law: philosophical foundations, deontics, and defeasible reasoning. In D. Gabbay, J. Horty, X. Parent, R. van der Meyden, & L. van der Torre, (Eds.), *Handbook of deontic logic and normative systems*, *Volume 2*, (pp. 657-764).

Hage, J. (2017). Legal Reasoning. In Hage, J., Waltermann, A., & Akkermans, B. (Eds.). (2017). *Introduction to law* (pp. 21-32). Springer.

Janssens, L., & Schaeken, W. (2016). 'But' Implicatures: A Study of the Effect of Working Memory and Argument Characteristics. *Frontiers in psychology*, *7*, 1520.

Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement learning: A survey. *Journal of artificial intelligence research*, *4*, 237-285.

Kahneman, D., & Tversky, A. (1982). The psychology of preferences. *Scientific American*, *246*(1), 160-173.

Katsos, N., & Bishop, D. V. M. (2011). Pragmatic tolerance: Implications for the acquisition of informativeness and implicature. *Cognition*, *120*(1), 67–81.

Marcus, R. B. (1966). Iterated deontic modalities. *Mind*, *75*(300), 580-582.

McNamara, P. (2004). Agential obligation as non-agential personal obligation plus agency. *Journal of Applied Logic*, *2*(1), 117-152.

McNamara, P. (2021). Logics for Supererogation and Allied Normative Concepts. In D. Gabbay, J. Horty, X. Parent, R. van der Meyden, & L. van der Torre, (Eds.), *The Handbook of Deontic Logic and Normative Systems*, *Volume 2* (pp. 155-306). College Publications.

Meinong, A. (1894). *Psychologisch-ethische untersuchungen zur werththeorie*. Leuschner & Lubensky.

Meyer, J. J., & Dignum, F. P. M. (1994). The paradoxes of deontic logic revisited: A computer science perspective (or: Should computer scientists be bothered by the concerns of philosophers?).

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial intelligence*, *267*, 1-38.

Noveck, I. A. (2001). When children are more logical than adults: experimental investigations of scalar implicature. *Cognition*, *78*(2), 165–188.

Noveck, I. A., & Posada, A. (2003). Characterizing the time course of an implicature: An evoked potentials study. *Brain and Language*, *85*(2), 203–210.

Nys, B. L., Wong, W., & Schaeken, W. (2024). Some scales require cognitive effort: A systematic review on the role of working memory in scalar implicature derivation. *Cognition*, *242*, 105623.

Pankratz, E., & van Tiel, B. (2021). The role of relevance for scalar diversity: a usage-based approach. *Language and Cognition*, *13*(4), 562-594.

Parent, X., & van der Torre, L. (2018). Dyadic Deontic Logic. In *Introduction to Deontic Logic and Normative Systems* (pp. 23–41). College Publications.

R Core Team (2023). _R: A Language and Environment for Statistical Computing_. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

Schaeken, W., Van Haeren, M., & Bambini, V. (2018). The Understanding of Scalar Implicatures in Children With Autism Spectrum Disorder: Dichotomized Responses to Violations of Informativeness. *Frontiers in Psychology*, *9*, 1266–1266.

van Tiel, B., Pankratz, E., & Sun, C. (2019). Scales and scalarity: Processing scalar inferences. *Journal of Memory and Language*, *105*, 93-107.

van Tiel, B., & Schaeken, W. (2017). Processing Conversational Implicatures: Alternatives and Counterfactual Reasoning. *Cognitive Science*, *41*(S5), 1119–1154.