

The Structure of Values and Norms

SVEN OVE HANSSON

*Royal Institute of Technology
Stockholm, Sweden*



CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York, NY 10011-4211, USA
10 Stamford Road, Oakleigh, VIC 3166, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

<http://www.cambridge.org>

© Sven Ove Hansson 2001

This book is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without
the written permission of Cambridge University Press.

First published 2001

Printed in the United States of America

Typeface Times Roman 10.25/13 pt. *System* QuarkXPress [BTS]

A catalog record for this book is available from the British Library.

Library of Congress Cataloging in Publication Data

Hansson, Sven Ove, 1951–

The structure of values and norms / Sven Ove Hansson.

p. cm. – (Cambridge studies in probability, induction, and decision theory)

Includes bibliographical references (p.) and indexes.

ISBN 0-521-79204-5

1. Values. 2. Norm (Philosophy) I. Title. II. Series.

BD232 .H297 2001

121'.8–dc21 00-063047

ISBN 0 521 79204 5 hardback

Contents

<i>Preface</i>	<i>page xi</i>
----------------	----------------

PROLEGOMENA

1. Formalization	3
1.1 Formalization and Idealization	3
1.2 The Virtues and Dangers of Formalization	5
1.3 Why Logic?	7
1.4 A Trade-Off Between Simplicity and Faithfulness	8
1.5 Formalizing Change	10

PART I: VALUES

2. Exclusionary Preferences	15
2.1 Some Basic Conditions	15
2.2 The Comparative Value Concepts	17
2.3 A Central Dilemma in Preference Logic	20
2.4 Minimal Principles for Choice-Guiding Preferences	23
2.5 Restrictability	26
2.6 Other Approaches to Rational Preferences	29
3. Preference States	33
3.1 Preference Sets	34
3.2 Preference Models	37
3.3 Intersectible Rationality Postulates	39
4. Changes in Exclusionary Preferences	42
4.1 The Basic Framework	43
4.2 Revision	46
4.3 Contraction	50
4.4 Subtraction	54
4.5 Addition	55

5. Constructing Combinative Preferences	57
5.1 Connecting the Two Levels	57
5.2 Representing Relata and Alternatives	60
5.3 Pairwise and Decision-Guiding Preferences	65
6. Pairwise Combinative Preferences	67
6.1 How to Compare Compatible Alternatives	67
6.2 From Exclusionary to Pairwise Preferences	70
6.3 Strict Preference and Indifference	73
6.4 Constructing the Representation Function	75
6.5 Transmitted Logical Properties	83
6.6 Superstructural Logical Properties	87
7. Decision-Guiding Combinative Preferences	94
7.1 Two Approaches to Decision Making	94
7.2 Weighted Preferences	97
7.3 Basic Criteria for Agnostic Preferences	100
7.4 Extremal Preferences	102
7.5 Further Conditions on Extremal Preferences	104
7.6 Properties of Extremal Preferences	107
7.7 A Résumé of Combinative Preference Relations	113
8. Monadic Value Predicates	114
8.1 Positive, Negative, and Circumscriptive Predicates	115
8.2 Negation-Comparing Predicates	117
8.3 Defining Good and Bad	118
8.4 Alternative Accounts of Good and Bad	122
8.5 Reducibility Theses	125

PART II: NORMS

9. A Starting Point for Deontic Logic	129
9.1 What Normative Predicates Refer To	129
9.2 The Multiplicity of Normative Predicates	131
9.3 Deontic Alternatives	134
9.4 Delineating Situationist Deontic Logic	137
10. Situationist Deontic Logic	139
10.1 Standard Deontic Logic (SDL)	139
10.2 Preference-Based Deontic Logic	143
10.3 The Logic of Positive and Contranegative Predicates	148
10.4 Representation Theorems	161
10.5 Canonical Deontic Predicates	164
10.6 Maximal Obeyable and Maxiconsistent Predicates	166

11. Conflicts and Counterfactuals	170
11.1 Shifting to Action-Guidance	170
11.2 Moral Dilemmas	173
11.3 Deontic Counterfactuals	175
11.4 Restrictive Conditionals	179
12. Rules and Normative Systems	182
12.1 Rules and Instantiations	182
12.2 Two Types of Application	185
12.3 Unrestrained Application	188
12.4 Consistent Application	193
12.5 Obeyable Application	200
12.6 Changing the Rules	203
12.7 Promulgation	203
12.8 Derogation	205
13. Legal Relations	208
13.1 Potestative Rules	208
13.2 Types of Legal Relations	212
13.3 Rights	214
13.4 A Comparison with Hohfeld's Typology	218
EPILOGUE	
14. Afterthought	225
14.1 Philosophical Relevance	225
14.2 Applications in Other Disciplines	226
14.3 A Difficult Trade-Off	228
<i>Proofs</i>	229
<i>References</i>	289
<i>Index of Symbols</i>	302
<i>General Index</i>	307

1

Formalization

Formalized philosophy is by no means uncontroversial. Obviously, this book would not have been written had not its author believed that useful insights can be gained from formal treatments of philosophical issues. The purpose of this chapter is to point out some of the advantages, but also some of the limitations, of formalization.

1.1 FORMALIZATION AND IDEALIZATION

A representation in formal language is always the outcome of a simplification for the sake of clarity, or in other words an *idealization*. To idealize in this sense means to perform a “deliberate simplifying of something complicated (a situation, a concept, etc.) with a view to achieving at least a partial understanding of that thing. It may involve a distortion of the original or it can simply mean a leaving aside of some components in a complex in order to focus the better on the remaining ones.”¹

Idealization – not necessarily in formal language – is omnipresent in science, and it seems to be so in philosophy as well. Many, probably most, of the crucial concepts in philosophical discourse originate through idealizations from nonphilosophical language.² As one example, it is common in moral philosophy to regard “John ought to

¹ McMullin 1985, p. 248. It is important to distinguish this sense of idealization from that of depicting something as better than it actually is. The two senses relate to different meanings of ‘ideal,’ namely on the one hand “[s]omething existing only as a mental conception” and on the other something that is “perfect or supremely excellent in its kind”. (OED) On the two senses of ‘ideal,’ see Hansson 1999d. On the role of idealization in philosophy, see also Hansson 1994.

² I prefer ‘nonphilosophical’ to the more common ‘prephilosophical,’ since there is no reason to believe that philosophy has, in its more than two and a half millennia of existence, been devoid of impact on thinking and language outside of philosophy.

...,” “It is a duty for John to ...,” and “John has an obligation to ...” as synonymous, in spite of the fact that there are occasions when common usage would accept one or two of these phrases but not the other(s).³ In this and similar cases, philosophers (tacitly) assume that there is, or can be constructed, a more fundamental and more straight-lined concept behind the embellished meanings of words and phrases in nonregimented natural language.

True, idealization is only one part of the transformation of elements from nonphilosophical language through which philosophical terminology is shaped. The construction of philosophical language also involves the creation of new distinctions and of terms that have no obvious counterparts in nonphilosophical language. Hence, philosophical terminology differs from nonspecialized language in two ways. First, it uses some words in different, idealized ways (e.g., ‘good,’ ‘value,’ ‘permission’). Second, it employs some linguistic innovations of its own (e.g., ‘consequentialism,’ ‘deontic’).

Some philosophers have wished to philosophize in “prephilosophical” language. In my view, this is an illusory undertaking, since nontrivial philosophical insights, with few exceptions, require more precision than what is immediately available in nonregimented language.⁴

In other academic disciplines, the relationship between specialized terminology and the nonspecialized concepts from which they originated may be fairly unproblematic. Physicists who theorize about heat or gravitation do not have to refer back to the nonphysicist’s concepts of warm and cold, or light and heavy, in order to justify their theoretical constructions. These scientific concepts have their own justifications, derived from experiments and other exact observations. Philosophers operating with concepts such as goodness or permission are not in this fortunate situation. These philosophical concepts have no justification apart from their capability of clarifying the corresponding nonphilosophical concepts. Hence, on one hand we have to deviate from the general-language meanings of our key terms in order to obtain the precision necessary for philosophical analysis; but on the other hand, if we deviate so far as to lose contact with general-language meanings, then the rationale for the whole undertaking will be lost.

³ See Section 9.2 and the references given there.

⁴ By ‘nontrivial,’ I mean nontrivial against the background of what has already been said by previous philosophers.

Since different idealizations can clarify different aspects of one and the same concept, it is futile to search for definite or uniquely correct philosophical analyses or explications. Different types and styles of idealizations of one and the same concept should be seen as complements rather than as competitors.

Therefore, a defence of formalized philosophy need not – and in my view could not reasonably – proceed by showing that formalization provides us with the one and only correct approach to philosophical problems. It is sufficient to show that some philosophical problems have aspects that can be clarified with formal methods.

Formalization in philosophy typically results from an idealization in two steps: first from common language to a regimented philosophical language and then from that regimented language into symbolic language.⁵ More often than not, most of the idealization takes place in the first of these steps. Therefore, what makes treatments in symbolic language special is not their distance to ordinary discourse (which can be surpassed by treatments in regimented natural language). Rather, what makes them special is the mathematical skill they require and the characteristic types of questions they give rise to.⁶

1.2 THE VIRTUES AND DANGERS OF FORMALIZATION

In the natural sciences, formalized theories have the preponderant advantage of being correlated with empirical measurements and thus being testable in a more exact way than informal theories. No such mensural correlation is available for formal theories in philosophy, such as theories of values and norms.⁷ Therefore, the claims that can

⁵ On some occasions, there may be more than two distinguishable steps of idealization. In a letter to the author (April 19, 2000), David Makinson pointed out that this is true of the formal representations of legal relations that are discussed in Chapter 13. In this case, there are three steps of idealization, passing through (1) ordinary language, (2) legal discourse, (3) Hohfeld's typology, and (4) formal representations of Hohfeld's distinctions.

⁶ It is interesting to note that the additive model used in utilitarianism, although controversial in other respects, seems to have escaped the negative reaction against mathematical representation that other formal models have encountered in some philosophical quarters. One reason for this may be that the mathematical skill it requires is so elementary. However, there is no reason why ease of mastering should be decisive for the accuracy of a formal model.

⁷ In this respect, formalized value theory can be compared to pre-Galilean physics. Many medieval physicists employed formal models of physical phenomena, but they did not use these models to predict the outcomes of measurements. See Livesey 1986 and the references given there.

be made for formalization in philosophy are weaker than those that can be made for formalization in the empirical sciences. In philosophy, the major virtue of formalization is the same as that of idealization in informal languages: Isolating important aspects helps to bring them to light. As compared to informal idealization, formalized treatments tend to be helpful in at least four more specific ways.

First, formalization incites definitional and deductive economy. It brings forth questions about the interdefinability of concepts and about minimizing the set of primitive principles of inference. In this book, formalized methods will be used to investigate the interdefinability of evaluative and normative concepts and also, within the first group, that between dyadic and monadic value concepts.

Second, formalization serves to make implicit assumptions visible. For example, in informal discourse on preferences it is often tacitly taken for granted that a well-defined set of alternatives (an alternative set) exists, consisting of the objects to be compared. In formal models of preferences, this assumption has to be made in a precise and explicit manner.

Third, formal theories can support delicate structures that would be much more difficult to uphold and handle in the less unambiguous setting of an informal language. Symbolic treatment has made it possible to penetrate some philosophical issues more deeply than would otherwise have been possible. One of the best examples of this is the relation between truth and language. It is difficult to see how Tarski's semantical analysis of the notion of truth could have been developed in a nonformalized setting. In this book, the discussion of different ways to derive preferences over parts from preferences over wholes draws heavily on distinctions that are readily available in formal languages but next to unattainable in nonformal language.

Fourth, formalization stimulates strivings for completeness. The rigorousness of a formal language is, for instance, necessary to make it meaningful to search for a complete list of valid principles of inference. Often enough, this search may uncover previously unnoticed philosophical problems. One example of this is the study of nonmonotonic inference. The introduction of rigorous formal notation (in particular, a nonmonotonic consequence relation) has led to much more thorough and extensive studies of patterns of nonmonotonic inference and of the relationships between these patterns.⁸ In this book, some of the results

⁸ Gabbay 1985; Makinson 1993.

on the properties of goodness and badness are examples of such strivings for completeness.

Obviously, formalization also has its dangers.⁹ In order to construct a workable formal model, the number of primitive notions has to be reduced to a very minimum. It is often cumbersome to include an additional factor into an already existing formal model. Partly for this reason, the philosophical logician runs the risk of becoming mentally locked in the world of one or a few formal models and therefore neglecting aspects of the real world not covered in these models. Another danger is spending too much time on problems that are mere artifacts of the formal model rather than on more general philosophical problems that the model can be helpful in elucidating. The so-called deontic paradoxes, which arise in certain models of deontic logic but not in informal normative discourse, are examples of this.

1.3 WHY LOGIC?

Formalization in philosophy is in practice virtually synonymous with formalization in logical language.¹⁰ Some of the pioneers of formal logic allotted to logic a unique status in philosophy. Bertrand Russell, for one, maintained that “every philosophical problem, when it is subjected to the necessary analysis and purification, is found either to be not really philosophical at all, or else to be, in the sense in which we are using the word, logical.”¹¹

Although this book conforms with philosophical tradition in employing logical formalism, I do not subscribe to such special claims for logic. There are no a priori grounds why logical languages should be better suited than other symbolic languages for modelling discourse on norms and values, nor, for that matter, for modelling real-world phenomena to which norms and values refer. The relative usefulness of logic is an open question, and treatments of the same subject matter in other types of formal languages should be welcomed.

⁹ Hansson 2000a.

¹⁰ There is no clear demarcation between logic and mathematics. Arguably, much if not most of mathematics can be reconstructed into some form of logic. Here, a logical language means a symbolic language of one of the types that are taught in logic courses and in textbooks on logic.

¹¹ Russell 1914, p. 14. Russell counted value theory among the “not really philosophical” topics, but his general view is compatible with the inclusion of value theory in the realms of logic. Cf. Davis 1966.

Against this background, it is not as problematic as some have thought to use truth-valued logic for modelling (subject matter expressed by) sentences that are not true or false.¹² A normative sentence such as

(1) Jane ought to help her brother.

cannot be true or false (or at least, for the sake of argument, let us assume that it cannot). Therefore, a formal sentence in a truth-valued language (such as $O\alpha$, where O stands for “ought” and α for “Jane helps her brother”) cannot, strictly speaking, represent sentence (1). It can, however, represent the following sentence:

(2) There is a valid norm to the effect that Jane ought to help her brother.

where validity is relative to some moral code or standard. We can assume that there is a one-to-one correspondence between sentences of the type represented by (1) and those of the type represented by (2). Truth and falsity are, of course, fully applicable to sentences of the latter type. Therefore, to the extent that a system of deontic logic adequately mirrors the properties of sentences such as (2), it also mirrors – somewhat more indirectly – those of sentences such as (1).¹³

1.4 A TRADE-OFF BETWEEN SIMPLICITY AND FAITHFULNESS

To formalize philosophical subject matter means to reduce it to a simplified formal model in order to get a clear view of some major aspects at the expense of others.¹⁴ Different formalizations may capture different features. As an example, probabilistic and nonprobabilistic theories of belief seem to capture different properties of human belief systems.

Philosophical or scientific model-making is always a trade-off between simplicity and faithfulness to the original. In philosophy, the subject matter is typically so complex that a reasonably simple model will have to leave out some philosophically relevant features. This makes it possible to devise a counterargument – typically in the

¹² See Makinson 1999 and the references given there.

¹³ Cf. Bengt Hansson 1969.

¹⁴ Cf. Merrill 1978, esp. pp. 305–311.

form of a counterexample – that seemingly invalidates the model. However, even if such a counterargument convincingly discloses an imperfection in the model, this is not necessarily a sufficient reason to give up the model. If the counterargument cannot be neutralized without substantial losses of simplicity, then an appropriate response may be to continue using the model, bearing in mind its weaknesses.¹⁵

As an example of this, it is assumed throughout this book that logically equivalent expressions can be substituted for each other. This assumption makes way for certain counterintuitive inferences, such as the *revenger's paradox*.¹⁶ Let p_1 , p_2 , and q be mutually exclusive expressions, such that p_1 or p_2 or q is logically true. Then

If Obligatory(not- p_1) then Obligatory(p_2 -or- q)

is logically true. Now let p_1 signify that John kills his wife's murderer, p_2 that he kills only persons other than his wife's murderer, and q that he kills nobody at all. It follows that if John ought not to kill his wife's murderer, then he ought to kill either only persons other than his wife's murderer, or no one at all.

Intersubstitutivity has similar effects on the logic of values. Let p denote that you receive €100 tomorrow, q that you receive €50 tomorrow, and r that you are robbed of all the money that you have the day after tomorrow. Presumably, you prefer p to q . By intersubstitutivity, you then also prefer $(p \& r) \vee (p \& \neg r)$ to q . However, the direct translation of $(p \& r) \vee (p \& \neg r)$ into natural language does not seem to be preferable to the direct translation of q into natural language. The disjunctive formulation of the comparison redirects attention, and the impression is created that each of the disjuncts is claimed to be preferred to q .

These and other counterintuitive inferences can only be avoided by giving up intersubstitutivity, thereby losing much of the simplicity and

¹⁵ This view is by no means uncontroversial among philosophers. Nicholas Rescher (1968, p. 238) maintains that a formal analysis of an informal concept "must, if acceptably executed, end with results that are fully compatible with ordinary conceptions." According to Hector-Neri Castañeda, the simpler of two theories should only be chosen if they account for exactly the same data; he thus gives faithfulness absolute priority over simplicity (Castañeda 1984, see in particular pp. 241–244).

¹⁶ The paradox was introduced in Hansson 1991a, to which the reader is referred for a somewhat more thorough discussion.

logical strength of the formal structure. It is, on balance, better for most purposes to endure the somewhat strange consequences of intersubstitutivity than to pay the high price for getting rid of them.¹⁷

1.5 FORMALIZING CHANGE

There are several ways in which formal logic can be used to express changes. In what may be called *time-indexed models*, a (discrete or continuous) variable is employed to represent time. The object of change (such as a state of the world, state of affairs, state of mind, state of belief, value state, etc.) can then be represented as a function of this variable, so that a state of the world (etc.) is assigned to each point in time. A further development of this framework is to make it non-deterministic by allowing for a bundle of functions, typically structured as a branching tree.

A quite different mode of representation is that of *input-assimilating models*.¹⁸ In such models, the object of change (such as a state of belief) is exposed to an input (such as a new piece of information) and is changed as a result of this. No explicit representation of time is included in models of this type. Instead, the characteristic mathematical constituent is a function that to each pair of a state and an input assigns a new state. (Nondeterminism can be achieved by replacing the function by a relation.)¹⁹ From the early 1980s and onward, input-assimilating models of belief states and databases have been the subject of a rapidly growing number of studies.²⁰

Input-assimilating models have the advantage of focussing on the causes and mechanisms of change. They are tailored to exhibit the effects of external causes on systems that are changed only in response to external causes (“inputs”) and are otherwise stable. This makes them extremely well suited to represent changes in most types of computerized systems, such as databases. It also makes them tolerably well suited to represent important aspects of changes in human states of mind. It is, at least for some purposes, a reasonable idealization to dis-

¹⁷ An interesting exception is the formal representation of free-choice permission. See Section 9.1.

¹⁸ Cf. Hansson 1999c, Section 1.3. In game theory, extensive form games may be seen as time-indexed and normal form games as input-assimilating.

¹⁹ Lindström and Rabinowicz 1989, 1991; Segerberg 1995.

²⁰ A seminal paper was Alchourrón, Gärdenfors, and Makinson 1985. For an overview, see Hansson 1999c.

regard those changes in a person's beliefs or values that have no immediate external cause in order to focus better on the mechanisms of externally caused changes.²¹ Input-assimilating models are also suitable to represent legal codes. At least ideally, a legal code remains the same unless a decision is taken to change it in a specified way.

In this book, input-assimilating models of changes in values and norms will be explored. However, there is no reason to believe that these are the only models that can shed light on the dynamics of values and norms. Other constructions, such as time-indexed models, may bring out important aspects that input-assimilating models hide from sight. The relative merits and demerits of different formal approaches can only be judged after each of them has been fairly thoroughly investigated.

²¹ Among the most important internally generated processes are those that aim at making a state of mind coherent or making it stable against further rational deliberations (reflective equilibrium). Recently, some attempts have been made to model such processes in otherwise input-assimilating models. See Hansson 1997b and Olsson 1997.