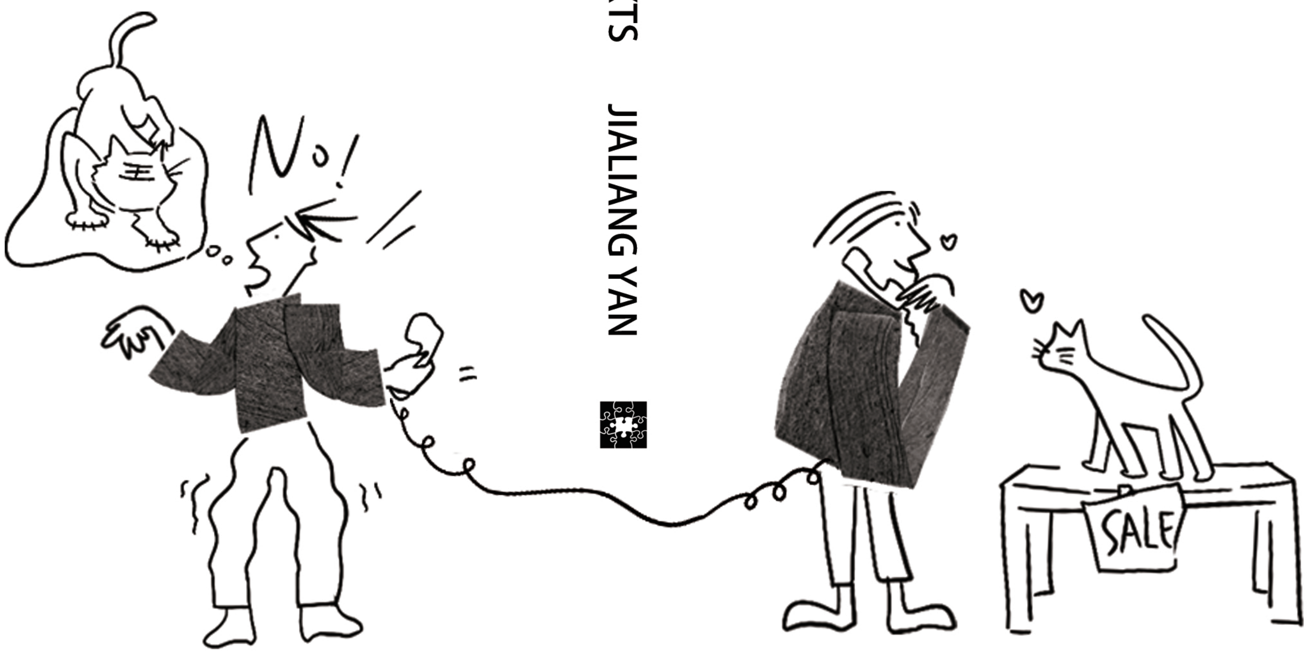
JIALIANG YAN

MONOTONICITY IN INTENSIONAL CONTEXTS

WEAKENING AND PRAGMATIC EFFECTS UNDER MODALS AND ATTITUDES



MONOTONICITY IN INTENSIONAL CONTEXTS JIALIANG YAN

MONOTONICITY IN
INTENSIONAL CONTEXTS

内涵语境中的单调推理

模态词和态度词下的弱化和语用效应

申请清华大学-阿姆斯特丹大学联合授予

博士学位论文

闫佳亮

二〇二三年九月

Monotonicity in Intensional Contexts

Weakening and Pragmatic Effects under Modals and Attitudes

Dissertation Submitted to

Tsinghua University and University of Amsterdam

in partial fulfillment of the requirement for a joint doctorate degree

by

Jialiang Yan

September, 2023

Monotonicity in Intensional Contexts

Jialiang Yan

Monotonicity in Intensional Contexts

ILLC Dissertation Series DS-2023-06

INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION

For further information about ILLC-publications, please contact

Institute for Logic, Language and Computation
Universiteit van Amsterdam
Science Park 107
1098 XG Amsterdam
phone: +31-20-525 6051
e-mail: illc@uva.nl
homepage: http://www.illc.uva.nl/

We acknowledge the generous support of a 1-year Chinese Scholarship Council (CSC) scholarship.

Cover design by Jialiang Yan and Bie Woo.
Printed and bound by Proefschriftenprinten.nl

ISBN: 978-90-833601-1-9

Monotonicity in Intensional Contexts
Weakening and Pragmatic Effects under Modals and Attitudes

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. ir. P.P.C.C. Verbeek
ten overstaan van een door het College voor Promoties ingestelde
commissie, in het openbaar te verdedigen in Room 329, Humanity Building,
Tsinghua University Campus (Beijing, China)
op maandag 11 september 2023, te 20.00 uur CST

door Jialiang Yan
geboren te Hebei

Promotiecommissie

Promotores:	dr. M.D. Aloni	Universiteit van Amsterdam
	prof. dr. F. Liu	Tsinghua University
Overige leden:	prof. dr. M.J.B. Stokhof	Universiteit van Amsterdam
	prof. dr. F.R. Smets	Universiteit van Amsterdam
	prof. dr. F. Roelofsen	Universiteit van Amsterdam
	prof. dr. J.F.A.K. van Benthem	Tsinghua University
	prof. dr. D. Zhu	Tsinghua University
	dr. M. Liu	Tsinghua University

Faculteit der Natuurwetenschappen, Wiskunde en Informatica

Dit proefschrift is tot stand gekomen binnen een samenwerkingsverband tussen de Universiteit van Amsterdam en Tsinghua University met als doel het behalen van een gezamenlijk doctoraat. Het proefschrift is voorbereid in de Faculteit der Natuurwetenschappen, Wiskunde en Informatica van de Universiteit van Amsterdam en in de School of Humanities van Tsinghua University.

This thesis was prepared within the partnership between the University of Amsterdam and Tsinghua University with the purpose of obtaining a joint doctorate degree. The thesis was prepared in the Faculty of Science at the University of Amsterdam and in the School of Humanities at Tsinghua University.

To the ones I love and who love me

your unwavering belief in me has fueled my journey.

CONTENTS

ACKNOWLEDGEMENTS

By virtue of good fortune, I have finally completed the writing of this doctoral thesis. This stroke of luck granted me the opportunity to study at Tsinghua University and the University of Amsterdam, introduced me to supervisors who provided unconditional support and help, caught the attention and mentoring of experts in my field, and allowed me to engage with like-minded friends for discussions.

I want to express my deep appreciation to my supervisors, Dr. Maria Aloni and Prof. Fenrong Liu, who co-supervised my thesis and significantly influenced my approach towards academic research. They respected all my academic interests and provided support to help me bring those interests to realization. They showed me how to do academic research with a sense of passion and commitment. I owe a great deal to Maria. Every moment spent working with her and seeking her advice was a pleasure. Maria has always respected my ideas, even when our views differed, she encouraged me to think deeper and helped me to strengthen my arguments. Her guidance has increased my confidence, which has played a big part in my academic development. I am deeply grateful to Fenrong, especially for the opportunity she gave me to participate in the joint degree program. Fenrong guided me in understanding the standards and norms of academia. No matter how busy she is, she always took my questions seriously, providing detailed answers to even the smallest concerns. I hold dear the time spent with both supervisors, filled with many memories of learning and growing for which I am thankful.

I would like to express my gratitude to the four Jin Yuelin Professors at Tsinghua, Prof. Johan van Benthem, Prof. Martin Stokhof, Prof. Dag Westerståhl, and Prof. Jeremy Seligman. Their help and support significantly contributed to the completion of my doctoral journey. I am particularly thankful to Johan. His guidance has been important to me, providing not only insightful suggestions for my research but also mentoring me in problem-solving and thinking. Special thanks go to Martin. Martin has guided and inspired me in the fields of semantics, philosophy, and logic. He has also taught me to balance academia with life. The genuine care from Martin and his wife, Karin Veraart, made my stay in Amsterdam relaxed, enjoyable, and truly unforgettable. I am grateful for Dag's suggestions on my thesis, particularly regarding the introduction part. His advice helped me navigate this challenging section. I appreciate Jeremy for all the knowledge

I've gained from his courses and our discussions. His advice on Chapter 3 of my thesis was crucial in refining and enhancing that work.

I am extremely grateful that my thesis has been reviewed by authoritative experts: Prof. Johan van Benthem, Dr. Mingming Liu, Prof. Floris Roelofsen, Prof. Sonja Smets, Prof. Martin Stokhof, Prof. Donghua Zhu. I sincerely appreciate them for agreeing to be in my thesis committee, dedicating their time to review my thesis and providing invaluable suggestions.

I would like to express my appreciation to my co-authors of the papers contained in the thesis: Dr. Maria Aloni, Prof. Fenrong Liu, and Dr. Kaibo Xie. Their contributions to the research are highly appreciated.

I would like to appreciate the other professors I met in classes or lectures. I was beneficial from their teachings and the interactions with them. My appreciation goes out to Dr. Alexandru Baltag, Dr. Dun Deng, Dr. Paul Dekker, Prof. Valentin Goranko, Dr. Sujata Ghosh, Prof. Xinwen Liu, Prof. Feng Qi, Prof. Robert van Rooij, Dr. Katrin Schulz, Dr. Chenwei Shi, Prof. Frank Veltman, Prof. Yanjing Wang and Dr. Fan Yang.

Special thanks go to all those at Tsinghua University and the University of Amsterdam who contributed to the success of the joint degree program. Their support and efforts have made it possible for this thesis to be presented and submitted at both institutions.

I would also like to thank the organizers of the Tsinghua Interdisciplinary Workshop on Logic, Language, and Meaning (TLLM). The topic of my PhD project came from the theme "Monotonicity in Logic and Language" at the second conference. Participation in this conference marked the start of my academic journey.

I am grateful for my friends. The discussions with them have been beneficial to me: Aleksi Anttila, Qian Chen, Fengxiang Cheng, Marco Degano, Alejandro Gracia Di Rienzo, Zhengyi Hong, Chen Ju, Tomasz Klochowicz, Soeren Knudstorp, Dazhu Liu, Dean McHugh, Zhiqiang Sun, Yunsong Wang, Fei Xue, Qingyu He and many others. A special thanks to Yiyan Wang and Lei Li, whose companionship made my doctoral journey happy and smooth.

I am grateful to my family, especially my parents, Bin Yan and Yanhong Liu. They are my warmest harbor and my firmest supporters. Their emphasis on education has been my greatest fortune. I also wish to thank my grandfather, Dekuan Liu, who is now in heaven. I regret that he could not see this thesis. I miss him very much, and I believe that if he could see this, he would be proud of me.

My thanks also go out to Raphael and Zitong for their assistance and care, which have brightened my life. I'm thankful to Bie for designing the cover and shaping visual beauty.

Lastly, I am thankful for every beautiful day that I have experienced.

Jialiang Yan

Beijing, July 2023

CHAPTER 1 INTRODUCTION

Chinese-style writing usually employs the term 缘 (yuán) 起 (qǐ) to denote the motivation for conducting research. This term derives from the core Buddhist concept of *Pratītyasamutpāda*, which implies the notion of dependent arising. I intend to use this term in this chapter instead of the English term “motivate”. This preference stems from the fact that 缘起 encapsulates not only the reasons for initiating research, but also acknowledges the role of serendipity and good fortune in academic research endeavors.

1.1 缘起 monotonicity

This dissertation aims to investigate issues that arise in the realm of modalities, with a particular focus on epistemic and desiderative modals. These issues are provoked by the presence of monotonic inferences in modal contexts. Consider the following line of reasoning:

(1) Baoyu has an allergy to seafood, with the exception of salmon. He enjoys consuming salmon.

a. Having salmon for dinner ⇒ Having seafood for dinner.

b. Baoyu wants to have salmon for dinner.

c. ? ⇒ Baoyu wants to have seafood for dinner.

While premise (1-b) holds true in this given context, the conclusion (1-c) appears infelicitous. It is not clear why Baoyu would want to eat seafood when he has an allergy to it. However, if monotonicity operates in this case, since having salmon implies having seafood, then (1-c) logically follows from (1-b), i.e., the desire for salmon implies the desire for seafood.

Analogous instances can be located in ancient Chinese literature. As an illustration, the subsequent exemplar is drawn from the *Xiaoqu* chapter of the Mohist Canons.[1]

(2) a. *Qí dì, měi rén yě* (其弟，美人也).

1 If one were to adopt the universal quantificational interpretation of “*ài měi rén*”, then this example could conceivably be construed as an instance of a failure in monotonic inference. For a more detailed analysis of this argument, please refer to Sun and Liu (2021).

Her younger brother is a handsome man.

b. *Ài dì, fēi ài měi rén yě* (爱弟，非爱美人也).

Loving her younger brother is not loving a handsome man.

What is the problem? Monotonic inferences encompass a variety of reasoning patterns pertaining to predicate replacement, such as substituting “seafood” with “salmon” in example (1), or “younger brother” with “handsome man” in example (2). Nevertheless, carrying out such predicate replacements can generate underspecification in modal contexts as in (1), or additional meaning as in (2), which yields unwarranted conclusions. For instance, in (1-b), only the desire for “salmon” is explicitly expressed; however, by employing monotonicity of the verb *want*, the conclusion in (1-c) possibly pertains to any type of seafood and implies that Baoyu desires seafood, no matter of which kind.

The monotonic aspects of modality semantics are seemingly challenged by the phenomena under consideration. Nevertheless, this dissertation proposes a different perspective: these phenomena, on my view, do not undermine the monotonicity of modalities. The central argument is that monotonicity with regards to modal expressions clashes with some aspects of language use. This tension gives rise to an erroneous perception of non-monotonicity in modality semantics.

Survey of related work. In the existing body of literature, the phenomena above have been already observed and discussed, e.g. Asher (1987); Heim (1992); Von Fintel (1999); Levinson (2003); Zimmermann (2006); Lassiter (2011); Crnič (2011); Cariani (2013); Von Fintel (2018); Jerzak (2019). A natural reaction to puzzles such as those presented in (1) or (2) is challenging the monotonic nature of the modal verb. Therefore, a straightforward solution could involve formulating a non-monotonic semantic analysis of the modal, and hence block the monotonic inferences.

A representative theory of the non-monotonicity approach is the analysis of *want* and other desire modals, established by Stalnaker (1984); Heim (1992). In Heim’s work, a semantic analysis of *want* is proposed which uses similarity relations and preference orderings. This analysis formalizes Stalnaker’s idea of desire attribution as a preference over doxasitic alternatives. In Heim’s semantics[1], *a wants* ϕ is true iff for every belief world w, every ϕ-world maximally similar to w is more desirable to the agent than any non-ϕ-world maximally similar to w. The semantics provides a non-monotonic interpretation of

1 Heim also formulated her proposal in a *context change potential* framework. For clarity and brevity, only the static version presented here. Further details can be found in Section 4 and Heim (1992).

want. The theory blocks the monotonic reasoning and predicts (1) to be invalid. The conclusion (1-c) is predicted to be false in the described scenario, as many Baoyu's doxastic worlds, wherein he has seafood for dinner, involve him eating crabs or other non-salmon seafood, which in turn results in a less desirable outcome compared to maximally similar worlds where he doesn't eat any seafood. However, the truth of (1-b) is still preserved. In those doxastic worlds where Baoyu has salmon, the outcome is more desirable compared to maximally similar worlds where salmon is not served in his meal.

Stalnaker-Heim semantics, along with other proposals such as Levinson (2003); Jerzak (2019), treat desire verbs as non-monotonic operators. Consequently, these frameworks interpret examples such as the one provided in (1), and others bearing analogous logical structures, such as Asher's puzzle (Asher (1987)), as instances of semantic failure. This interpretation emphasizes the role of non-monotonicity as a key factor in explaining the logical breakdown observed in these examples.

Another famous example of a possible monotonicity failure discussed in the literature is embodied in Ross' paradox arising in deontic domains, as delineated in Ross (1944). Ross' paradox seems *prima facie* to present a different challenge from the previously mentioned examples, thereby adding another layer of complexity to our understanding of these puzzling phenomena. Ross, in his seminal work, observed that certain imperatives, such as "post the letter", do not imply further imperatives, such as "post the letter or burn it". Analogously, this gives rise to similar intuitions when applied to deontic modal expressions of permission or obligation.

(3) a. Posting the letter ⇒ posting the letter or burning it.
    b. Baoyu may (ought to) post the letter.
    c. Baoyu may (ought to) post the letter or burn it.

The drawn inference in (3) seems invalid, as the conclusion may suggest the permissibility of burning the letter - an implication not explicitly stated in the initial premise (3-b). The conclusion (3-c) seems to imply that Baoyu can choose between the two actions, while only one action is permitted by the premise (3-b). The conclusion (3-c) is an example of a *free choice* permission (see Von Wright (1968); Kamp (1974)).

A plausible approach to Ross' paradox suggests that the inference is intuitively invalid because the second action introduced in the conclusion, namely burning the letter, is impermissible in this scenario. Several works, including Lewis (1979); Sayre-McCord

(1986); Cariani (2013); Lassiter (2011), have pursued this intuitive path, which blocks the monotonic inferences. For instance, Lewis (1979) provides a seminal formal analysis of deontic expressions. Lewis postulates that when an agent is given a permission to execute a particular action, it does not thereby allow the outcome to occur in any manner the agent deems fit. It suggests that in a communication, granting permission may broaden an agent's action scope, especially when they were previously uncertain of their abilities to perform such actions. However, this expansion does not necessarily include all correlated actions. To illustrate, when an individual is permitted to post the letter, this authorization does not automatically imply that she can post a letter in a flaming state. This highlights the inherently non-monotonic nature of permissions: the permission to post the letter does not imply the permission to post it or burn it. In Lewis' analysis this follows from the *performativity* of permissions (see Ninan (2005); Portner (2009)). Consequently, there should exist regulations that govern the potential expansion given by permissions. Yablo (2010) has put forth a semantics with the aim to formalize the principles underlying this implicit understanding of permissions.

In alignment with the analysis of desire discussed above, these theories propose a non-monotonic account of deontic modals. These approaches, albeit in different domains, desire and deontic modals, offer semantic strategies and attempt to resolve the puzzles by blocking monotonicity inferences. I will argue that these semantic approaches fall short in two significant respects.

Firstly, theories of non-monotonicity, which are fundamentally semantic-based, may encounter challenges in addressing all variations of monotonicity puzzles. These theories often face difficulties when attempting to explain certain phenomena, such as free choice inferences, which are generally considered to be triggered by pragmatic effects. For example, the Stalnaker-Heim semantics and Lewis' approach may not provide a full solution for puzzles that share a similar structure with Ross' paradox. These theories struggle to offer a cogent explanation for the emergence of free choice inferences. Furthermore, as highlighted by von Fintel in Von Fintel (2012), if deontic and desire modalities are not upward monotonic, then NOT OUGHT or NOT WANT should not licence negative polarity items (NPIs) according to the standard theory of NPIs, as the environment would not be downward-entailing.

Secondly, these theories basically depend on the unique features of a particular modal, such as the preference structure of *want* or the performativity of permission. Therefore, their effectiveness is bounded to a specific domain, and they lack the versatility needed for application in different modal contexts. For example, they may not fully address puzzles in epistemic contexts or with epistemic modals that differ significantly in semantic structure from deontic and desire modals.

In order to address the first issue, von Fintel employs a pragmatic strategy (as discussed in his works Von Fintel (1999, 2012, 2018)). Within his theoretical framework, the deontic or desire modalities have monotonic semantics, while pragmatic explanations account for the associated complexities. Consider, for example, the following scenario discussed by Von Fintel (1999). In this situation, Baoyu exhibits no desire to buy a couch at its original price; however, if it were available at a 25% discount, he would buy it.

(4) a. Buying the couch at a 25% discount ⇒ Buying the couch.
    b. Baoyu wants to buy the couch at a 25% discount.
    c. Baoyu wants to buy the couch.

(5) Baoyu doesn't want to buy the couch.

In von Fintel's theory, the statement (4-c) logically follows from (4-b), as his semantics refine the Hintikka-style semantics for *want* and function monotonically. If all desirable worlds involve buying the couch at a 25% discount, all these worlds are buying-couch worlds. However, (5) also can be true, which seems paradoxical. This can be reconciled by understanding that (4-c) and (5) are interpreted w.r.t different modal bases. For (4-c), the modal base expands due to (4-b), introducing 25%-discounted-couch worlds. Yet for (5), no such expansion occurs, leading there are only original price worlds in Baoyu's doxastic set. Thus Baoyu believes that the couch's original price is his only option, which explains his reluctance to buy.

Therefore in this theory the perceived infelicity is attributed to a change in the contextual belief, leading to a shift in the modal base. Yet, this pragmatic explanation appears to be less compelling when applied to the instance outlined in (1). This stems from the fact that the allergies remains constant, irrespective of any alterations in beliefs. Additionally, his theory, in isolation, cannot account for Ross' paradox. This necessitates the incorporation of a specialized theory of the free choice phenomenon. We will delve into a detailed examination of his theory in Section 4.

In addition, it is worth noting that von Fintel's pragmatic explanations are intrinsically tied to the inherent features of the modal, such as the relation of desire ascription to belief. Consequently, von Fintel's approach cannot entirely disentangle itself from the dual concerns mentioned above.

Weakening effect triggered by monotonicity. As discussed, the two respects, previously outlined, continue to persist and pose challenges to both semantic and pragmatic solutions to monotonicity puzzles, thereby leaving room for further investigation the issue:

- Is it possible to address different monotonicity-related puzzles (e.g., Asher's puzzle and Ross' paradox) across different modal domains in a uniform way?

In order to address it, this dissertation puts forth a general and uniform methodology for examining monotonicity in intensional contexts. It is postulated that the proposed theoretical framework will provide an explanation for a diverse range of puzzles across various domains. This constitutes the first theme of the dissertation, namely, a framework study concerning the issue of monotonicity.

In this dissertation, the primary argument is that a weakening effect, triggered by monotonicity under attitudes, is the principal cause of the phenomena and puzzles under investigation which are labeled as the *weakening effect triggered by monotonicity* (henceforth WEM). To illustrate this, consider again example (1), which is repeated below in (6).

(6) Baoyu has an allergy to seafood, with the exception of salmon. He enjoys consuming salmon.

 a. Having salmon for dinner ⇒ Having seafood for dinner.
 b. Baoyu wants to have salmon for dinner.
 c. ? ⇒ Baoyu wants to have seafood for dinner.

The conclusion in (6-c) is weaker than the premise in (6-b) since a model satisfying (6-c) can contain more possibilities (or bigger) than the one satisfying (6-b), such that the models illustrated in Figure 1.1.

The fundamental assumption is that the weakening resulting from applications of monotonicity gives rise to pragmatic effects, similar to the free choice and ignorance inferences triggered by disjunctive statements. The weakening could result in the audience interpreting the conclusions drawn from monotonicity by the bigger models, thereby leading to a possible pragmatic failure. I will further propose that in the absence of explicit disjunctive statements, as in (6), such effects arise because the relevant predicates can be

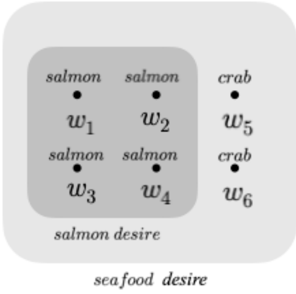


Figure 1.1 Salmon desire and seafood desire

reinterpreted disjunctively. In example (6), we argue that the predicate "seafood" can be reinterpreted as "salmon or non-salmon seafood". Consequently, the sentence in (6-c) is rephrased as "Baoyu wants salmon or non-salmon seafood for dinner." which pragmatically implies that "It is ok for Baoyu to have non-salmon seafood for dinner." And this implicature can be seen as the rationale behind rejecting (6-c).

Contrasting with von Fintel's analysis, our WEM theory also involves model enlargement, but its impetus differs. Whereas von Fintel's theory attributes enlargement to the premise, ours is triggered by monotonicity. Chapter 4 provides a more formal analysis and supporting arguments of WEM theory.

The theory of WEM shows a versatile applicability. Pragmatic effects of weakening are not inherent semantic characteristics of a specific modality or the consequence of a modal context, but rather represent a persistent tension between logical reasoning and language use. This tension remains constant and is not eliminated by altering the modal domain. Consequently, the WEM theory is applicable across a diverse range of modal settings. Additionally, by reinterpretation, we can find the connection between monotonicity puzzles and certain pragmatic effects, so the WEM theory can also be employed to address distinct puzzle types.

In Chapter 2, 3, and 5, we show the application of the proposed framework by employing it to analyze a variety of distinct puzzles in epistemic and desire contexts. These puzzles are associated with different modalities. This leads us to the second theme of the dissertation: an exploration of various modalities within the theoretical framework of WEM. This framework enables a more detailed observation of interactions between modalities, and promotes a thorough investigation into their logical and semantic properties.

Mapping the research landscape. This thesis explores then two tiers of inquiry through the lens of WEM:

- Framework level: the thesis aims to develop the WEM theory and explores its underlying mechanisms and principles. By doing so, it strives to enhance our understanding of the nature of monotonicity reasoning and its relationship with modal expressions.
- Implementation level: the thesis investigates how WEM interacts with different modalities. Specifically, it studies the influence of WEM on the understanding and interpretation of various modal expressions and attitude verbs in natural language.

The thesis focuses on epistemic contexts to address the implementation level issue, e.g., investigating the interaction between the non-classical properties of epistemic modalities and monotonicity. Additionally, a desiderative context is employed to address the framework level issue. Through an analysis of paradoxes related to monotonicity, a theory of WEM is proposed. The approach in this thesis is logic-based, and also using insights from linguistics, philosophy, and cognition.

This research contributes a new perspective to understand and bridge logic and language, as well as semantics and pragmatics. Firstly, the study examines the logical characterization of reasoning in natural language concerning monotonicity. It becomes evident through the observation of WEM that inconsistencies or mismatches may arise between the meaning conveyed by such reasoning and the logical predictions generated. The question arises whether these discrepancies result from the inherent ambiguity of natural language or if the formalization of logic is insufficient to capture the diverse range of linguistic expressions.

By investigating WEM and the modalities under WEM theory, we discover the importance of further examination of the mediation of logic and language. Namely the factors that enable language usage laws to be determined and formalized by logic. For instance, the concept of weakening may be one of such mediation. Examining weakening allows us to investigate the proper formalization of generalizations and abstractions in natural language, thereby enhancing our understanding of the intricate relationship between logic and language.

On the other hand, monotonicity is typically considered as a semantic property. However, in the dissertation it has been observed that this form of semantic reasoning can generate pragmatic effects, which in turn influence our understanding of the semantic aspects of modalities. This appears to establish a cycle of reciprocal impact. Consequently, the precise borderline between pragmatic and semantic reasoning emerges as a pertinent inquiry. Furthermore, it is essential to determine whether the reasoning instigated by monotonicity reasoning should be classified as semantic or pragmatic. Within the logical framework of BSML (Aloni (2022)), which has been employed in the present dissertation, such reasoning is triggered by pragmatic effects, but can be semantically characterized. Although this issue has not been extensively investigated, the observed **WEB** phenomena suggest that the distinction between semantic and pragmatic reasoning is not as rigid as previously assumed.

Thus far, we have outlined two primary topics under investigation: firstly, the examination of a generally theoretical framework for WEM; and secondly, the exploration of specific modalities under WEM theory. In the next section, a presentation of the thesis's structural organization and the interrelations between its constituent elements will be provided.

1.2 The structure of the dissertation

This thesis comprises five research papers, which have either been published or are currently under review.

Chapter 2. In Chapter 2, we investigate what it means to know or believe that something might be the case. The chapter addresses the issue focusing on the epistemic possibility expressed by English *might* when embedded under the propositional attitude verbs *know* and *believe*. We present some puzzles related to WEM to highlight the challenges arising from such *know-might* and *believe-might* sentences. We propose a framework to solve the puzzles, in which epistemic *might* is defined as quantifying over the epistemic possibilities in an information state, and belief is formalized in term of a plausibility ordering. In contrast to classical epistemic logic, the factivity of knowledge is treated as a presupposition rather than being solely dependent on the reflexivity of the accessibility relation. Based on these, a team-based modal logic BSEL is established, which is treated as an epistemic variant of Aloni's (2022) BSML.

- Chapter 2 was written by J. Yan and M. Aloni, and is an extended version of the conference paper J. Yan and M. Aloni. Monotonicity under Knowledge: Epistemic Possibility in Logic and Conversations. *The 11th Scandinavian Logic Symposium (SLSS 2022)*. Bergen, Norway, 2022. The extended version: Knowing and Believing an Epistemic Possibility now is under review.

- Contribution: J. Yan initiated the paper with formal framework and arguments which were then further developed and extended in collaboration with Aloni.

Chapter 3. This chapter investigates properties of epistemic *might* and its interaction with knowledge and belief in multi-agent contexts, diverging from traditional single-agent studies. Epistemic might is argued to be sensitive to an agent's perspective. Therefore, epistemic *might*-claims should be interpreted in relation to an agent's epistemic state. As a result, the same *might* claim could be evaluated differently by different agents. We propose a framework to formalize this phenomenon, integrating a team-based semantics of BSEL and a two-dimensional semantics for Epistemic Friendship Logic (Seligman et al. (2011, 2013)). This enables us to employ a mechanism of perspective shifting to model the perspective sensitivity of epistemic *might*. Additionally, we demonstrate that this framework can also serve as a valuable tool for explaining the phenomenon of faultless disagreement in a systematic and comprehensive manner.

- Chapter 3 is based on the paper: J. Yan, M. Aloni and F. Liu. Shifting Perspectives: An Multi-agent Extension of BSEL. *Manuscript*. The paper now is ready for submission.
- Contribution: J. Yan extended the framework BSEL proposed in Chapter 2 by integrating the Epistemic Friendship Logic (Seligman et al. (2013)) and BSML (Aloni (2022)) frameworks. It was further developed and extended in collaboration with the co-authors.

Chapter 4. This chapter studies the theory of WEM in the context of desire. It presents three typical puzzles that challenge monotonic semantics for the verb *want*. The chapter advances the argument that all three puzzles can be reduced to the cases of Free Choice (FC) inferences. To explain this, a reinterpretation theory is proposed, in which predicates can be reinterpreted as disjunctive statements that lead to pragmatic effects. In order to show this, a theory of reinterpretation is put forth. Based on this theory, predicates can be reinterpreted as disjunctive statements, thereby triggering specific pragmatic effects. The chapter also reviews Aloni's proposal QBSML and adds a reinterpretation function. The explanation of the three puzzles provids an account of the WEM that arise from semantic inferences (monotonicity reasoning) and lead to pragmatic inferences (free choice inferences or ignorance effects). Finally, the chapter concludes with a discussion of potential applications to deontic cases.

- Chapter 4 is based on two papers. One is an extension of J. Yan. The Overtone of Monotonicity under Desire. *Proceedings of the ESSLLI 2022 Student Session*. Published online, 2022. Another is J. Yan and F. Liu. Monotonic Opaqueness in Deontic Contexts. In Proceedings of 5th AWPL, Liao B., Wáng Y. (eds) *Context, Conflict and Reasoning*. *Logic in Asia: Studia Logica Library*. Springer, Singapore. 2020.
- Contribution: J. Yan proposed a pragmatic account for the puzzles of monotonicity under desire and a reinterpretation theory with broad applicability.

Chapter 5. After arguing that monotonic reasoning does not affect the semantics of *want* in Chapter 4, this chapter proposes an interpretation of desire by taking into account causal inferences. To formalize this idea, a desire-causality model is constructed by combining the betterness model in preference logic and the causal model in the logic for causal reasoning. Subsequently, a logic for desire based on this semantics is developed, and an axiomatization for the formal system is presented.

- Chapter 5 is based on the paper of K. Xie and J. Yan. A Logic for Desire Based on Causal Inference. *Journal of Logic and Computation*. Accepted (extended journal version of the LORI21 conference paper).
- Contribution: J. Yan contributed to the development of the model's fundamental concept and the formal analysis, as well as contributing to part of the writing. The axiomatization was undertaken by K. Xie.

Although each of the four chapters represents an independent research article, they are interrelated and intrinsically linked. Together, they constitute the core of this dissertation and systematically address the implementation-level and framework-level questions raised in the preceding section.

The interdependence of the four chapters is illustrated in Figure 1.2, with Chapter 2 and 3 focusing on epistemic modals, and Chapter 4 and 5 examining desiderative modals. Chapter 2 addresses implementation level concerns by investigating the properties of epistemic modals under the tension created by WEM. As a follow-up research, Chapter 3 examines epistemic *might* in a multi-agent environment.

On the other hand, Chapter 4 focuses on framework level issues by analyzing the paradoxes of monotonicity and distinguishing the semantic and pragmatic aspects of WEM. The baseline framework to account for WEM is QBSML (Aloni and van Ormondt (2023)),

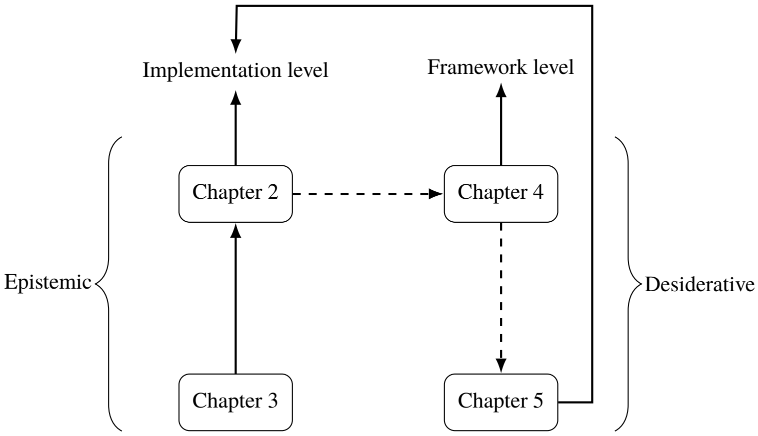


Figure 1.2 The relations between the chapters

which is a first-order version of BSML. In the sense of the continuity of the formal framework, Chapter 4 indirectly follows Chapter 2. Although Chapter 5 is methodologically independent of the other three chapters as it examines desire based on preference logic and causal inferences instead of state-based logic, it continues the views of Chapter 4. Since the analysis in Chapter 4 shows that the apparent failures of monotonic reasoning does not affect the semantic properties of the verb *want*, Chapter 5 does not consider related monotonicity puzzles when studying the semantics of the *want*. Instead, it focuses on the semantic characteristics of the notion of desire, such as whether it is related to preference and causality, and these characteristics determine whether its semantics should be monotonic or not.

The research on modalities presented in this dissertation has either a direct or an indirect connection to the WEM. As such, it necessitates a comprehensive understanding of the effect, which is predicated on a foundational grasp of monotonicity as a concept. In the next section, we shall provide definitions of monotonicity as a preliminary foundation for the ensuing discussions and analyses.

1.3 Preliminaries

In this section, we provide precise definitions of monotonicity reasoning at both the propositional and quantified levels. It is important to note that the monotonicity reasoning referenced in the following chapters adheres to these definitions.

1.3.1 Monotonic inferences in propositional level

Monotonic inferences are ubiquitous in natural language. The monotonicity principle is a key to comprehending the mechanisms of human reasoning. In Aristotelian logic (especially for syllogisms) or in the ancient Chinese Mohist logical thought, many valid patterns of reasoning can be understood in terms of monotonicity explicitly or implicitly (Eijck (2005); Sun and Liu (2021)). As such, researchers have investigated and formalized the inferences in natural language through the lens of monotonicity. For example, Generalized Quantifier Theory (GQT, Barwise and Cooper (1981); Van Benthem (1984); Keenan and Westerståhl (1996); Peters and Westerstahl (2006)) examines quantifiers in terms of their monotonicity, which is treated as a general concept that characterizes order-preserving attributes over partially ordered domains. Additionally, the field of Natural Logic (NL, Van Benthem (1988, 1995); Valencia (1991); Moss (1987, 2015); Icard (2012)) has proposed a Monotonicity Calculus program that marks expressions according to their monotonicity properties, and applies this calculus in a proof system. It also has broad applications in computational linguistics (e.g.MacCartney (2009)). Furthermore, monotonicity plays a critical role in the analysis of Negative Polarity Items (NPIs, Fauconnier (1975, 1979); Ladusaw (1979); Von Fintel (1999)). The semantic scope of negation, which corresponds to downward-monotonic environments, is considered to be the NPI-licensing environment.

As argued in Icard (2012), it is not easy to get a handle on what monotonicity actually is. A plausible explanation takes monotonicity to describes a functional property of "ordering preserving".

Given a set of possible worlds W, a propositional operator over W is a mapping from propositions to truth values: $\mathcal{P}(W) \rightarrow \{0, 1\}$. Monotonicity of a propositional operator can be defined based on whether it reverses or preserves the direction of entailment in its propositional argument.

Definition 1 (Monotonicity of propositional operators)**:** *For a unary propositional operator Δ,*

- *Δ is upward monotonic ($\Uparrow$) iff if Δp and $p \rightarrow q$ then Δq, for arbitrary propositions p and q.*
- *Δ is downward monotonic ($\Downarrow$) iff if Δq and $p \rightarrow q$ then Δp for arbitrary propositions p and q.*
- *Δ is non-monotonic (-) iff Δ is neither upward nor downward monotonic.*

Monotonicity reasoning in modal logic can be characterized by the following inference rule:

$$\frac{\varphi \rightarrow \psi}{\Delta\varphi \rightarrow \Delta\psi}$$

1.3.2 Monotonic inferences in quantified level

To set the scenes, we first introduce monotonicity of determiners based on Peters and Westerstahl (2006).

Definition 2 (Monotonicity of quantifiers)**:**

An upward monotonic quantifier of type $< t_1, ..., t_m >$ *in ith argument is defined w.r.t a model* M *as:*

- *If* $Q_M[A_1, ..., A_i ..., A_m]$ *and* $A_i \subseteq A'_i \subseteq M^{n_m}$*; then* $Q_M[A_1, ..., A'_i, ..., A_m]$*, where* $1 \leq i \leq k$.

A downward monotonic quantifier of type $< t_1, ..., t_m >$ *in ith argument is defined as:*

- *If* $Q_M[A_1, ..., A_i, ..., A_m]$ *and* $A'_i \subseteq A_i \subseteq M^{n_m}$*; then* $Q_M[A_1, ..., A'_i ..., A_m]$*, where* $1 \leq i \leq k$.

To understand the above definitions, we can consider the following examples including (generalized) quantifiers of type $< 1, 1 >$:

(7) a. Everyone is running. So every boy is running.
    b. Everyone is running fast. So everyone is running.

(8) a. Some dogs are black. So some animals are black.
    b. Some dogs are black. So some animals are colored.

(9) a. None of my family members smokes. So none of my parents smokes.
    b. None of my family members smokes. So none of my family members smokes on the sly.

(10) a. Not all girls have long hair. So not all people have long hair.
     b. Not all girls have long hair. So not all girls have long curly blonde hair.

(11) a. Most adults are well educated. So most adults are educated.

In examples (7) to (10), we can reason monotonically on both noun phrases and verb phrases. In other words, the quantifiers in these examples have double monotonicity: *ev-*

ery is $\downarrow [every] \uparrow$[1]; *some* is $\uparrow [some] \uparrow$; *no* is $\downarrow [no] \downarrow$; *not all* is $\uparrow [not\ all] \downarrow$. These four quantifiers correspond to the traditional Aristotle's square of opposition that we are familiar with:

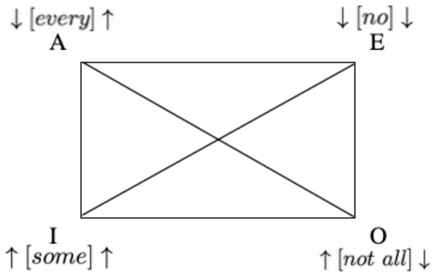


Figure 1.3 Square of opposition about monotonicity.

More research on the square regarding monotonicity can be found in Westerståhl (2012). When we move to generalized quantifiers, things get a bit complex. Some generalized quantifiers are monotonic, e.g., *most* in the sentence (5), while some are not, e.g., *even* and *exactly two*.

The monotonic step which guarantees logical validity could be named **Predicate Replacement** which amounts to the substitution of a specific (general) predicate with a general (specific) one in the upward (downward) context generated by the upward (downward) monotonic determiner. Therefore, based on Van Benthem and Liu (2020), the patterns of monotonicity reasoning can be defined in the first-order logic as following:

Definition 3 (Mon-inf)**:** [2]

- *Upward monotonicity reasoning*

  From $\phi(P)$ *and* $\forall x(P(x) \to Q(x))$*, it follows that* $\phi(Q/P)$*, where* $\phi(Q/P)$ *is the result of replacing each occurrence of* P *in* ϕ *by* Q*.*

- *Downward monotonicity reasoning*

  From $\phi(Q)$ *and* $\forall x(P(x) \to Q(x))$*, it follows that* $\phi(P/Q)$*, where* $\phi(P/Q)$ *is the result of replacing each occurrence of* Q *in* ϕ *by* P*.*

In conclusion, this section has provided an introduction to definitions of monotonic-

1 The left arrow symbolizes the monotonic inferences of the first argument of the determiner in the sentences, while the right arrow signifies the monotonic inferences of the second argument. Additionally, the upward arrow indicates upward monotonicity, and the downward arrow represents downward monotonicity.

2 In Van Benthem and Liu (2020), there is another notion of monotonicity, that is *Extension Increase*, and it can be defined as Mon-sem (only focus on upward case):

$M \equiv_P^+ M'$, where M' is the same model of M except for the Interpretation of P and $I(P) \subseteq I'(P)$, then $M', v \vDash \phi(P)$.

ity. As we advance to the primary subject matter of the dissertation, we shall extend our understanding of this fundamental concept in a broader logical framework.

CHAPTER 2 KNOWING AND BELIEVING AN EPISTEMIC POSSIBILITY

Knowledge, belief and their relation have received considerable attention in philosophy and logic, e.g. Hintikka (1962); Williamson (2000); Stalnaker (2006); Baltag et al. (2014); Shi et al. (2018). Much attention has been devoted to determining what is knowledgeable and what it is to know or believe that something is the case – a fact. But what does it mean to know or believe that something might be the case? What does it mean to know or believe an epistemic possibility? In this chapter we will address this issue by focusing on the epistemic possibility expressed by English *might* when embedded under the propositional attitude verbs *know* and *believe* as in (1)[1].

(1) a. Daiyu believes that it might be raining.
    b. Daiyu knows that it might be raining.

In addition, in this chapter, we will show the significant role that the WEM plays in examining the interaction among epistemic *might*, *know* and *believe*. To illustrate this point, we will present a puzzle that shows how the WEM serves as an indicator explaining the tension between the semantic and pragmatic attributes of given modalities.

2.1 Introduction

Epistemic *might* has a number of non-classical properties. For example, it has been argued to give rise to so-called **epistemic contradictions** (see Wittgenstein (1953); DeRose (1991); Veltman et al. (1996); Yalcin (2007); Mandelkern (2019) among others). It is infelicitous to simultaneously assert that p is the case and that p might not be the case, as evidenced by the oddness of the following sentence.

(2) # It is raining and it might be not raining.

1 The distribution of *might*-sentences is limited – not all attitude verbs can have a *might*-sentence as a complement. For example, as shown in Hacquard and Wellwood (2012), *might* cannot embed under desideratives.

(i) # Daiyu wants/prefers that it might be raining.

Epistemic *might* also generates **free choice effects** (see Zimmermann (2000); Veltman (1996); Von Fintel and Gillies (2008); Hawke and Steinert-Threlkeld (2021); Aloni (2022) etc.), as illustrated in (3), where a conjunctive meaning can be drawn from a modal disjunctive statement.

(3) a. Daiyu might be in Suzhou or Jinling.
b. $\rightsquigarrow$ Daiyu might be in Suzhou and Daiyu might be in Jinling.

Both properties – infelicity of epistemic contradiction and the free choice behaviour – are preserved when *might*-sentences are embedded under *know* or *believe* (see Zimmermann (2000); Yalcin (2007); Beddor and Goldstein (2018) etc.).

(4) # Daiyu believes/knows that it is raining and it might be not raining.

(5) Baoyu believes/knows that Daiyu might be in Suzhou or in Jinling. $\rightsquigarrow$ Baoyu believes/knows that Daiyu might be in Suzhou and he believes/knows that Daiyu might be in Jinling.

The non-classical behaviour of *might*, which as illustrated in (4) and (5) is preserved under embedding, leads to tensions with well-known properties of knowledge and belief. In this chapter, we will introduce some puzzles highlighting these tensions.

First, the verbs *know* and *believe* are normally assumed to be **upward monotonic**, as evidenced by the following inference.

(6) Daiyu believes/knows that Baoyu is angry and crying. $\Rightarrow$ Daiyu believes/knows that Baoyu is angry.

The first challenge arises from the combination of the monotonicity of *know*/*believe* and the free choice behaviour of *might*. This tension will be illustrated in our first puzzle where, applying free choice principles, we will derive epistemic contradictions from plausible assumptions.

A second challenge is generated by the contrast between knowledge and belief. Knowledge is normally taken to differ from belief. The norms of knowledge are much stronger than those of belief. The attitude of knowing requires full confidence in the truth of a proposition while the attitude of believing is less demanding. As a consequence, it is normally assumed that **knowledge entails belief**. Does this contrast apply to the *know*-

might and *believe-might* sentences as well? If so, will there be entailment from *know-might* to *believe-might* sentences? We will present a second puzzle addressing these issues which will challenge the common assumption of entailment from knowledge to belief at least when we consider embedded *might*-sentences.

And finally consider the **factivity of knowledge**. The verb *know*, as a factive predicate, implies that its clausal complement is true. Hence, a *know-might* sentence should also imply the truth of the embedded *might*-sentence. However, a *might*-sentence does not always express an objective truth. Building on non-classical semantics of *might* motivated by examples like (2) (see Veltman et al. (1996); Gillies (2001); Yalcin (2007); Willer (2013) etc.), in our logic *might*-formulas will be associated with assertability conditions rather than truth-conditions (Aloni (2022); Hawke and Steinert-Threlkeld (2021)). The use of *know* rather than *believe* will then imply a commitment to the assertability (rather than the truth) of the complement clause on the part of both the speaker and the subject (Lasersohn (2009)). In the last part of the article we will address a technical issue arising for a formalisation of this idea using a reflexive accessibility relation and solve it by modelling factivity as a presupposition (Spector and Egré (2015)).

The structure of this chapter is organized as follows. In Section 2.2 & 2.3, we present our two main puzzles. In Section 2.4, we provide an account of epistemic free choice inferences as neglect-zero effects, presenting an epistemic variation of Bilateral State-based Modal Logic (BSML, Aloni (2022)) where epistemic *might* is treated as an operator directly quantifying over the epistemic possibilities in a state. This will give us a strategy to solve puzzle 1. In Section 2.5, we further define knowledge and belief and their contrast. We use plausibility models building on (Van Benthem (2007); Baltag and Smets (2016)) to define belief. The classical **S5** modality is used to represent the informational modality in our proposal. We define the notion of knowledge in terms of presupposition and informational modality. In this way we can capture the factive inferences of *know-might* formulas and the projection of the factivity presupposition under negation. This will give us a strategy to solve puzzle 2. In Section 2.6 we present BSEL, an extension of BSML, where all these ingredients are put together and we present main results and applications. The distinction between the usage of *know* and *believe* in natural language will be discussed in Section 2.8. Finally, Section 2.9 concludes.

2.2 Puzzle 1: a tension between monotonicity, factivity and free choice

In this section, we present the first puzzle. We observe that applications of monotonicity, factivity of knowledge and epistemic free choice principles create a tension that leads to paradoxical conclusions. We start by introducing the three properties mentioned above, and then present the puzzle.

2.2.1 Preliminaries

2.2.1.1 Monotonicity under knowledge

In natural language, monotonicity is reflected in the semantic properties of various operators. In this chapter, we focus on propositional operators like knowledge or belief. The semantics of the verbs *know* and *believe* is commonly assumed to be upward monotonic. In Section 1.3.1, we have already defined monotonicity of propositional operators. We now present some examples that demonstrate the monotonic inferences under the attitude verbs *know* and *believe*.

(7) a. Every dog barks. ⇒ Fido barks.
 b. Daiyu knows/believes that every dog barks. ⇒ Daiyu knows/believes that Fido barks. (⇑)

(8) a. The apple is red. ⇒ The apple is red or green.
 b. Daiyu knows/believes that the apple is red. ⇒ Daiyu knows/believes that the apple is red or green. (⇑)

A basic feature of monotonicity is that negation will reverse its inferential pattern. So upward monotonicity under negation will be reversed to downward monotonicity.

(9) a. Daiyu doesn't know/believe that Fido barks. ⇒ Daiyu doesn't know/believe that every dogs bark. (⇓)
 b. Daiyu doesn't know/believe that the apple is red or green. ⇒ Daiyu doesn't know/believe that the apple is red. (⇓)

2.2.1.2 Factivity of Knowledge

The verb *know* is a factive verb. Knowing p requires the truth of p.

Definition 4 (Factivity of knowledge)**:**

An agent α knows a proposition p only if p.

Formally: $\texttt{KNOW}\ p \vDash p$

This principle is among the least controversial in epistemology. It originates in Plato's *Meno* and *Theaetetus*, in which Plato characterised knowledge as "justified true belief". While the justification and belief conditions, as the Achilles' heel of Plato's analysis, have been repeatedly challenged and attacked in the post-Gettier era, epistemologists generally agree that what is false cannot be known[1], see e.g. Armstrong (1973); Davidson (2001); Kelp and Pritchard (2009); Mitova (2018). While knowledge is factive, there is no such requirement for belief, formally $\texttt{BEL}\ p \nvDash p$.

2.2.1.3 Free choice and ignorance inferences under epistemic modals

Free choice triggered by disjunctive statements Von Wright (1967) observed that in natural language disjunctive permission statements can entail the permission of both disjuncts. This phenomenon is generally labeled as *free choice* (Kamp (1974), see Meyer (2020) for an overview). The same conjunctive inference can be drawn from necessary disjunctions, possible disjunctions and even from wide scope disjunctions, as summarised in the following:

- $\Box$-free choice
  $\Box(p \vee q) \vDash \Diamond p \wedge \Diamond q$
- $\Diamond$-free choice
  $\Diamond(p \vee q) \vDash \Diamond p \wedge \Diamond q$
- Wide scope (disjunction) free choice
  $\Diamond p \vee \Diamond q \vDash \Diamond p \wedge \Diamond q$

Free choice effects can be generated in both deontic and epistemic contexts, for example:

(10) $\Box$-free choice

a. Daiyu ought to study poetry or learn embroidery. $\rightsquigarrow$ Daiyu is permitted to study poetry and she is permitted to learn embroidery. (Deontic)

b. That must be Daiyu or Baoyu. $\rightsquigarrow$ That might be Daiyu and that might be

1 Some scholars argued that there exists false knowledge and people can know what is approximately wrong, see Hazlett (2010); Turri (2011); Bricker (2018); Buckwalter and Turri (2020).

Baoyu. (Epistemic)

(11) $\Diamond$-free choice

a. You may go to the beach or to the cinema. $\rightsquigarrow$ You may go to the beach and you may go to the cinema. (Deontic Kamp (1974))

b. Mr. X might be in Victoria or in Brixton. $\rightsquigarrow$ Mr. X might be in Victoria and Mr. X might be in Brixton. (Epistemic Zimmermann (2000))

(12) Wide scope free choice

a. Detectives may go by bus or they may go by boat. $\rightsquigarrow$ Detectives may go by bus and detectives may go by boat. (Deontic Zimmermann (2000))

b. Daiyu might be in Suzhou or she might be in Jinling. $\rightsquigarrow$ Daiyu might be in Suzhou and Daiyu might be in Jinling. (Epistemic)

Free choice effects are not derived by the rules of classical modal logic where the standard semantics for modals is combined with a Boolean analysis of disjunction. Providing a logical account of free choice is not trivial. For example, plainly adding free choice inferential patterns as axioms to classical modal logic will not do. For we would be able to derive from any possibility any other possibility. There are various approaches to explain the puzzle of free choice in the linguistic debate, including semantic accounts (see Aloni (2007); Zimmermann (2000); Geurts (2005); Simons (2005) etc.) and pragmatic accounts (Vainikka (1987); Fox (2007); Chemla (2009); Alonso-Ovalle (2006) etc.).

Ignorance inferences out of disjunctions Also a plain disjunction can give rise to conjunctive modal inferences. These are called modal disjunction inferences, see Gazdar (1979); Zimmermann (2000); Büring (2008); Mendia (2015) among others.

- Modal disjunction or ignorance inference[1]

  $p \vee q \models \Diamond p \wedge \Diamond q$

Here is an example from natural language:

(13) a. Baoyu is in Rongguofu or in Ningguofu.

b. $\rightsquigarrow$ Baoyu might be in Rongguofu and he might be in Ningguofu.

Zimmermann (2000) further observed that when a plain disjunction is embedded un-

1 The inference $\Diamond p \wedge \Diamond q$ in combination with the scalar inference $\neg(p \wedge q)$, give rise to proper ignorance effects: $\neg\Box p \wedge \neg\Box q$.

der a propositional attitude verb such as *believe* and *know*, it still gives rise to an ignorance effect, as illustrated in (14).

(14) a. Daiyu believes/knows that Baoyu is in Rongguofu or in Ningguofu.

b. $\rightsquigarrow$ Daiyu believes/knows that Baoyu might be in Rongguofu and that he might be in Ningguofu.

In addition we observe that also free choice effects are generated under *believe* or *know*:

(15) a. Daiyu believes/knows that Baoyu might be in Rongguofu or in Ningguofu.

b. $\rightsquigarrow$ Daiyu believes/knows that Baoyu might be in Rongguofu and that he might be in Ningguofu.

So both ignorance and free choice inferences triggered by disjunctive statements are preserved under knowledge and belief.

- Ignorance inferences under knowledge and belief
  $\texttt{KNOW}\ (p \vee q) \vDash \texttt{KNOW}\ \Diamond p \wedge \texttt{KNOW}\ \Diamond q$
  $\texttt{BEL}\ (p \vee q) \vDash \texttt{BEL}\ \Diamond p \wedge \texttt{BEL}\ \Diamond q$
- Free choice inferences under knowledge and belief
  $\texttt{KNOW}\ \Diamond(p \vee q) \vDash \texttt{KNOW}\ \Diamond p \wedge \texttt{KNOW}\ \Diamond q$
  $\texttt{BEL}\ \Diamond(p \vee q) \vDash \texttt{BEL}\ \Diamond p \wedge \texttt{BEL}\ \Diamond q$

After having introduced the relevant principles we can now present our first puzzle highlighting a tension among them.

2.2.2 Puzzle 1

Suppose that Daiyu knows the Netherlands is in Europe. From this innocent premise we will derive the epistemically contradictory conclusion that the Netherlands is in Europe and the Netherlands might be in Asia. The steps of the argument can be summarized as follows:

Puzzle 1 (version 1)**:**

(16) a. Daiyu knows that the Netherlands is in Europe. *(Assumption)*

b. Daiyu knows that the Netherlands is in Europe or Asia. *(Monotonicity)*

c. Daiyu knows that the Netherlands might be in Asia. *(Ignorance effect)*

d. *The Netherlands might be in Asia.* *(Factivity of Knowledge, c)*

e. *The Netherlands is in Europe.* *(Factivity of Knowledge, a)*

Given the observation in (15), we present also another version of the puzzle, in which free choice effects are generated rather than ignorance inferences.

Puzzle 1 (version 2)

(17) a. *The Netherlands is in Europe.* *(Assumption)*

b. *Daiyu knows that the Netherlands might be in Europe.* *(Assumption)*

c. *Daiyu knows that the Netherlands might be in Europe or Asia.* *(Monotonicity)*

d. *Daiyu knows that the Netherlands might be in Asia.* *(Free choice effect)*

e. *The Netherlands might be in Asia.* *(Factivity of Knowledge, d)*

Every step in (16) and (17) is supported by the principles we have introduced above. Let's just focus on (16). The case of (17) runs in a parallel fashion, except (17-a) is here an independent assumption rather than a derived conclusion. From the innocent assumption (16-a), upward monotonicity of the verb *know* is applied, together with the rule of *disjunction introduction* from classical logic. Then we derive a disjunction under *know*. (16-c) is drawn by the ignorance effect. Finally, by factivity of knowledge we derive both "the Netherlands might be in Asia and the Netherlands is in Europe", leading to an epistemic contradiction:

(18) # The Netherlands is in Europe, but it might be in Asia (not in Europe).

Sentence (18) has the same form as (19), which we discussed in the introduction.

(19) # It is raining, but it might not be raining.

In logic-based accounts we can capture the infelicity of these sentences by validating the following principle:

$$\phi \wedge \Diamond \neg \phi \vDash \bot$$

As argued in Mandelkern (2019), a general reason to consider (19) as contradictory lies in the fact that the utterance "it might not be raining." proposes to keep open the possibility that it is not raining. Conversely, the utterance "it's raining." proposes to rule out this possibility. The conjunction of these two statements would entail making both proposals, which cannot be executed simultaneously. Therefore, in (18), the established fact that "the Netherlands is in Europe." negates the possibility of the Netherlands not being in Europe, as implied by the latter part of (18). This results in a paradox. It's important to note that epistemic contradictions are not classical contradictions, for instance, the negation of an epistemic contradiction is not a tautology.

Incidentally, another significant feature of epistemic *might* is non-factive:

$$\Diamond\phi \nvDash \phi$$

This is more intuitive and obvious. The epistemic possibility doesn't provide a factual assertion about the world.

An adequate analysis of epistemic *might* should derive epistemic contradictions and also preserve non-factivity. But there is a tension between these two desiderata. If the epistemic contradiction is logically contradictory, then it follows that $\Diamond\phi$ entails ϕ which contradicts non-factivity.

This tension has motivated non-classical or so-called expressivist analysis of epistemic *might*. Expressivists argued that *might*-sentences are not evaluated in term of truth conditions in possible worlds but rather in terms of assertability with respect to the information state of the relevant agent (see Veltman et al. (1996); Yalcin (2007); Schroeder (2008); Hawke and Steinert-Threlkeld (2021) etc). A proposition p is taken to be assertable by an agent if the agent approves a body of information that supports p. More discussions on assertability can be found in Schroeder (2008). We will take this proposal as our baseline analysis for *might*. It also motivates us to adopt a state-based semantics to capture the assertability, and we will explain this later.

Back to Puzzle 1. It seems that the most straightforward solution to this puzzle is to block the inferences (16-c) or (16-d). There are two possible ways of blocking (16-c): getting rid of the upward monotonicity of knowledge, or getting rid of the ignorance principle (or free choice). Either way would, of course, naturally block (16-d). In addition, as a third option, if we embrace the statement (16-c), then we can simply break away from the factivity of knowledge to block (16-d). All these strategies, however, are not very appealing. For they require us to abandon principles that have a firm intuitive and philosophical basis. We will therefore dismiss these strategies and propose instead a different solution to the puzzle. The aim of our approach is to preserve all discussed principles, some of which will be semantic and others will be pragmatic, and provide an accurate semantics of epistemic *might* to capture its non-classical properties. Before presenting our solution, we will introduce our second puzzle.

2.3 Puzzle 2: the entailment from *know-might to believe-might*

In this section, we present the second puzzle which is relative to the so-called *Epistemic Entailment Thesis* (EET) which assumes that knowledge entails belief. We first present cases of epistemic contradictions in the scope of *believe* and observe their infelicity. We then observe that these sentences can be rescued if we substitute *believe* with *know* in one of its occurrences. We ascribe this phenomenon to a contrast between *know-might* and *believe-might* sentences.

2.3.1 Believing an epistemic contradiction

The verb *believe* is normally taken to be non-factive. It merely expresses the agent's positive attitude towards a proposition. What is believed should be compatible with the doxastic state of the agent. So it is incoherent to believe any contradiction including epistemic contradictions. For example:

(20) # Daiyu believes that it is raining and it is not raining.

(21) # Daiyu believes that it is raining and it might be not raining[1] (Beddor and Goldstein (2018)).

The infelicity persists also if we let *believe* take narrow scope[2]:

(22) # Daiyu believes that it is raining but she believes that it might not be raining.

The proposition "Daiyu believes that it might not be raining" means that the possibility of not raining is implied by Daiyu's belief, so "it is not raining" is compatible with the doxastic state of Daiyu. Whereas "Daiyu believes that it is raining" entails that the proposition "it is raining" holds over Daiyu's doxastic states and excludes the possibility of not raining in his belief. This attributes to Daiyu an incoherent belief. Therefore, it seems to be inadmissible to assert (22).

Now, notice that, as observed by Hawthorne et al. (2016); Mandelkern (2019) among others, if we replace *believe* with *know* in the latter clause of (22) then the result seems to be coherent:

1 We focus on the static analysis on the epistemic *might* so there are same predictions of $\neg p \wedge \Diamond p$ and $p \wedge \Diamond \neg p$.

2 Beddor and Goldstein (2018) argued that it is not infelicitous to assert statements like (22). It shows an indisputable fact that humans are fallible but still rational, so sometimes it is coherent to believe ϕ and also believe $\Diamond\phi$. While in this chapter, we argue that it is inconsistent to assert (22). We think that the process of drawing the incoherent belief can be rational, but the statements like (22) as a conclusion is not. We will leave this issue to the future discussion.

(23) Daiyu believes that it is raining but she knows that it might not be raining.

This sentence seems felicitous. On the one hand, belief is weak and people can hold wrong beliefs. So even though Daiyu knows it might not be raining, she can still have a belief that it is raining outside. On the other hand, such belief is likely to be merely one piece of evidential information (e.g., she thought she heard the sound of rain), which is not sufficient to be evidence to doubt or challenge Daiyu's knowledge. It doesn't rule out all possibilities of not raining. And in this case, Daiyu's knowledge only needs that there is at least one possibility of not raining compatible with her information state. So, in either way, her belief of raining will not be incoherent with Daiyu's knowledge.

However, the acceptability of (23) is puzzling if we assume that knowledge entails belief (see Mandelkern (2019)). We explain this in Puzzle 2.

2.3.2 Puzzle 2

If knowledge entails belief, then if we accept (24-a) then we should accept (24-c) which is inconsistent.

Puzzle 2:

(24) *a.* *Daiyu believes that it is raining but she knows that it might not be raining.*

b. *Daiyu knows that it might not be raining ⇒ Daiyu believes that it might not be raining.*

c. *# Daiyu believes that it is raining but she believes that it might not be raining.*

Hawthorne et al. (2016) argues that we should not take (24-c) to be inconsistent. They think that the consistency of (24-a) is evidence that such form of combination is not epistemically defective. However, we observe that there is a strong contrast in acceptability between (24-a) and (24-c), a contrast we would like to account for.

Puzzle 2 highlights a problematic aspect of a widely accepted assumption in epistemology, i.e., the epistemic entailment thesis. This principle *prima facie* seems to be correct. As introduced in Section 2.2, knowledge is traditionally defined as "justified true belief", so belief seems to be automatically derived from knowledge[1]. In this chapter however, motivated by the case illustrated in Puzzle 2, we will embrace a restricted version of the entailment thesis. We propose that the entailment from knowledge to belief

1 As always in philosophy, this principle has been challenged, some subtle counterexamples can be found in Radford (1966); Myers-Schulz and Schwitzgebel (2013) etc.

only applies to factual propositions and should be blocked in the case of *know-might* and *believe-might*.

We argue that this failure of entailment is because of the contrast between *know-might* and *believe-might* sentences. First, belief is weak. So in pragmatics, the verb *believe* is weaker than *know* in the sense that *know* presupposes the truth or assertability of its complement, while *believe* does not. In other words, the usage of *know* involves the norm of assertion which is what the verb *believe* lacks. Besides, *know-might* sentences explicitly involve the related speaker's knowledge of *might*-sentences. *Believe-might* sentences instead involve the speaker's knowledge only in an implicit way. For *believe-might* sentences convey that the speaker does not agree with the subject of the assertion otherwise she would have used *know* instead of *believe*.

On the other hand, in semantics, an expressivist analysis suggests that epistemic possibilities are relative to assertibility rather than truth. And *know* is a factive verb. So we argue that *know-might* sentences imply the assertability of the *might*-sentences with respect to the information of the related speaker. Whereas *believe* does not have the same restrictions on complements. *Might*-sentences embedded under *believe* are not required to be asserted, in agreement with the non-factivity of belief.

Therefore our strategy for Puzzle 2 is to block the entailment from *know-might* sentences to *believe-might* sentences. But for other cases, we preserve EET.

2.4 Free choice and ignorance inferences as neglect-zero effects

After having presented the two puzzles, we now discuss the strategies to solve them. In this section, we present the main ingredients of our analysis of puzzle 1, which include a non-classical semantics of epistemic *might* implemented in the framework of Bilateral State-based Modal Logic (BSML, Aloni (2022)).

Puzzle 1 involves principles governing knowledge and epistemic possibilities that have been discussed in logic, philosophy and linguistics. We believe that all these principles are correct and mutually consistent. We just need an appropriate and refined method of bringing them together. In this section, we will propose our account of Puzzle 1 with this intention.

First we introduce BSML. One of the advantages of BSML is that it derives both semantic and pragmatic inferences. So in addition to monotonic inferences, which are derived as semantic entailments, we can also derive free choice and ignorance inferences as pragmatic effects. Following the expressivist tradition, we further propose a non-classical account of *might* as a quantifier over epistemic possibilities in an information state. By assuming that all formulas are evaluated with respect to information states rather than to single worlds, BSML will allow us to solve the tension between the infelicity of epistemic contradictions and the non-factivity of *might*.

2.4.1 Bilateral State-based Modal Logic

As mentioned in Section 2.2.1.3, free choice and ignorance inferences triggered by disjunctive statements can be analyzed in term of pragmatics. (Neo-)Gricean approaches provide a possible explanation which appeals to some conversational reasoning, in which semantics and pragmatics are modeled as two separate components. Such (Neo-)Gricean approaches however typically cannot account for embedded cases of free choice inferences or cases of wide scope free choice (see Aloni (2022); Zimmermann (2000) for more arguments).

In Aloni (2022), Aloni proposed a formal account of free choice inferences in Bilateral State-based Modal Logic. On this account, free choice and ignorance inferences are not the result of some conversational reasoning but rather a consequence of a cognitive tendency operative in ordinary conversation. Aloni called this tendency *neglect-zero*. On this hypothesis, people when interpreting a sentence construct representations of the world and in doing so they systematically neglect models that validate the sentence by virtue of some empty configuration – *zero models*. From a cognitive perspective, people are assumed to prefer imagining a concrete reality rather than an abstract one, in the process of understanding language. Zero-models are models which verify a sentence by virtue of some empty set. The following example provides a straightforward illustration.

Statement: **At least one apple is in each bowl**

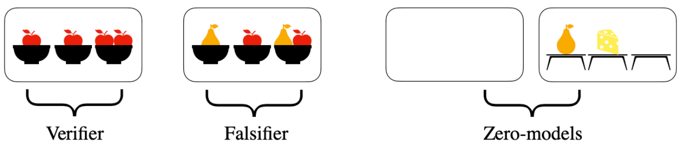


Figure 2.1 Examples of zero-models

The first picture shows a model which verifies the statement, and the second shows a model which falsifies it. The last two pictures represent zero-models which verify the statement by virtue of an empty witness set. Since there are no bowls in the model the statement is vacuously verified.

To model the neglect zero tendency, Aloni uses tools from team semantics (see Väänänen (2007); Yang and Väänänen (2017)) where formulas are interpreted with respect to a set of points of evaluation (i.e. a team) rather than single ones as in classical modal logic.

Assume a Kripke model $M =< W, R, V >$

- Classical modal logic: $M, w \vDash \phi$, $w \in W$
- Team-based modal logic: $M, t \vDash \phi$, $t \subseteq W$

In BSML, a team is interpreted as an information state s. So formulas in BSML are evaluated w.r.t information states. Moreover BSML is bilateral in the sense that the assertability and rejectability of a sentence are modeled by *support* ($\vDash$) and *anti-support* ($\Dashv$) conditions respectively rather than truth.

- $M, s \vDash \phi$ means ϕ is assertable in s where $s \subseteq W$
- $M, s \Dashv \phi$ means ϕ is rejectable in s where $s \subseteq W$

BSML further employs the non-emptiness atom (NE) from team logic (Yang and Väänänen (2017)). The atom NE is supported by a state s iff the state is non-empty.

$$M, s \vDash \mathrm{NE} \;\text{ iff }\; s \neq \varnothing$$

Then a *pragmatic enrichment function* (denoted by $[\cdot]^+$) can be defined in terms of a systematic intrusion of NE in the process of interpretation (Aloni (2022)).

Definition 5 (Language $\mathcal{L}$ of BSML)**:** *Let A be a set of sentential atoms $A = \{p, q, r, ...\}$.*

$$\phi := p \mid \neg\phi \mid \phi \wedge \phi \mid \phi \vee \phi \mid \Diamond\phi \mid \mathrm{NE}$$

where $p \in A$

Definition 6 (Pragmatic enrichment)**:** *A pragmatic enrichment function is a mapping $[\cdot]^+$ from the NE-free fragment of language $\mathcal{L}$ of BSML to $\mathcal{L}$ such that*[1]*:*

- $[p]^+ = p \wedge \mathrm{NE}$
- $[\neg\phi]^+ = \neg[\phi]^+ \wedge \mathrm{NE}$
- $[\phi \vee \psi]^+ = ([\phi]^+ \vee [\psi]^+) \wedge \mathrm{NE}$

1 As argued in Aloni (2022), pragmatic enrichments modeled in BSML result from speakers' tendency to avoid "zero-models" when creating pictures of the world representations during sentence interpretation. It will be suspended when doing In logico- mathematical reasonings in which many inferences rely on the availability of zero-models. However, the pragmatic enrichments in logical-mathematical reasoning, the neglect-zero effect is set aside. This is because many inferences in this type of reasoning rely on the availability of zero-models.

- $[\phi \wedge \psi]^+ = ([\phi]^+ \wedge [\psi]^+) \wedge \mathrm{NE}$
- $[\Diamond\phi]^+ = \Diamond[\phi]^+ \wedge \mathrm{NE}$

As argued in Aloni (2022), pragmatic enrichments modeled in BSML result from speakers' tendency to avoid "zero-models" when creating pictures of the world representations during sentence interpretation. However, in logical-mathematical reasoning the pragmatic enrichments are globally set aside, namely the neglect-zero effects are globally suspended. This is because many inferences in this type of reasoning rely on the availability of zero-models. Consider, for instance, the Addition Principle denoted as $p \vDash p \vee q$. Aloni further categorizes BSML into BSML$^{\varnothing}$, BSML^{lex}, and BSML*, each serving to formalize the conditions under which neglect-zero enrichments are globally suspended, locally existent, and existent in a manner that is both global and non-detachable, respectively. For a more detailed discussion, refer to Aloni (2022).

The following result can be derived in BSML.

Fact 1:

$$M, s \vDash [\phi]^+ \Rightarrow M, s \vDash \phi$$
$$M, s \dashv [\phi]^+ \Rightarrow M, s \dashv \phi$$

In fact, the atom NE also connects to the conversational principle "avoid contradictions". In state-based semantics, the empty state vacuously supports every classical formula, including contradictions. The interpretation of NE requires the supporting state to be non-empty, and hence, it captures the idea that the empty state which represents abstract elements (zero or the empty set in mathematics) and supports logical absurdity is excluded in the process of natural language interpretation. Therefore, the pragmatically enriched formulas, which rule out the empty state, are used to model human's reasoning in conversations.

In BSML, free choice effects can be directly derived for pragmatically enriched formulas. In what follows, we show how this works.

BSML adopts the standard notion of disjunction from team semantics, which is sometimes called *split* or *tensor* disjunction. A split disjunction $\phi \vee \psi$ is supported in a state s iff s is a union of two substates each supporting one of the disjuncts.

$$M, s \vDash \phi \vee \psi \text{ iff } \textit{there are } t, t' : t \cup t' = s \text{ and } M, t \vDash \phi \text{ and } M, t' \vDash \psi.$$

It is easy to see that, the evaluation of a pragmatically enriched disjunction $[a \vee b]^+$ is very different from that of the plain formula $a \vee b$. The former is assertable in s iff s can be split into two **non-empty** substates, one supporting a and another supporting b. The following figure illustrates this distinction.

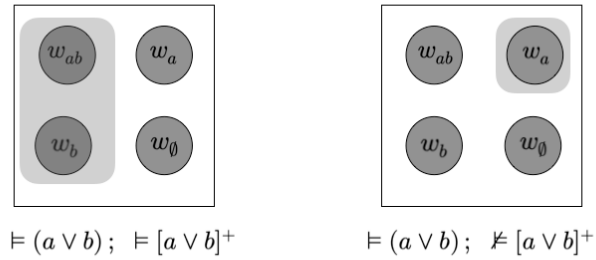


Figure 2.2 Comparison of $(a \vee b)$ and $[a \vee b]^+$

The light grey area stands for information state s. Based on the semantics of split disjunction, the first model verifies both $a \vee b$ and $[a \vee b]^+$. Whereas, in the second model, the information state s is a single world w_a which rejects the enriched disjunction $[a \vee b]^+$. For a pragmatically enriched disjunction requires both disjuncts to be live possibilities, and there is no world verifying b in this state. But this model verifies $a \vee b$. In fact, the second model is a zero model for $a \vee b$, as b is verified by virtue of an empty witness. The core effect of neglect-zero is to disallow such zero-models when evaluating the enriched formulas.

A model for $\mathcal{L}$ is a triple $M = \langle W, R, V \rangle$, where W is a set of possible worlds. R is an accessibility relation on W. V is a world-dependent valuation function: $A \times W \rightarrow \{0, 1\}$. Formulas of $\mathcal{L}$ are interpreted in models M with respect to an information state $s \subseteq W$. Both *support* ($\vDash$) and *anti-support* ($\Dashv$) conditions are specified. The semantic clauses are defined as follows (Aloni (2022)).

Definition 7 (Semantic clauses)**:**

$M, s \vDash p$	*iff*	$\forall w \in s : V(w, p) = 1$
$M, s \Dashv p$	*iff*	$\forall w \in s : V(w, p) = 0$
$M, s \vDash \neg\phi$	*iff*	$M, s \Dashv \phi$
$M, s \Dashv \neg\phi$	*iff*	$M, s \vDash \phi$
$M, s \vDash \phi \vee \psi$	*iff*	$\exists t, t' : t \cup t' = s$ *and* $M, t \vDash \phi$ *and* $M, t' \vDash \psi$
$M, s \Dashv \phi \vee \psi$	*iff*	$M, s \Dashv \phi$ *and* $M, s \Dashv \psi$
$M, s \vDash \phi \wedge \psi$	*iff*	$M, s \vDash \phi$ *and* $M, s \vDash \psi$
$M, s \Dashv \phi \wedge \psi$	*iff*	$\exists t, t' : t \cup t' = s$ *and* $M, t \Dashv \phi$ *and* $M, t' \Dashv \psi$
$M, s \vDash \Diamond\phi$	*iff*	$\forall w \in s : \exists t \subseteq R[w] : t \neq \emptyset$ *and* $M, t \vDash \phi$
$M, s \Dashv \Diamond\phi$	*iff*	$\forall w \in s : M, R[w] \Dashv \phi$
$M, s \vDash \mathrm{NE}$	*iff*	$s \neq \emptyset$

$M, s \dashv \mathrm{NE}$ *iff* $s = \varnothing$

where $R[w] = \{v \in W \mid Rwv\}$

A proposition p is supported by the information state s iff p is true in every worlds in s; and p is anti-supported by s iff it is false in every worlds in s. A negation $\neg\phi$ is supported by the state s iff s anti-supports ϕ, and $\neg\phi$ is anti-supported iff s supports ϕ. As said, a disjunction is supported by s iff s is a union of two substates each supporting one of the disjuncts. And $\phi \vee \psi$ is anti-supported iff both ϕ and ψ are anti-supported by s. A conjunction $\phi \wedge \psi$ is supported by s iff each conjunct is supported by s, and $\phi \wedge \psi$ is anti-supported by s iff s is a union of t and t' and each substate anti-supports one of conjuncts. The operator $\Diamond$ is interpreted in term of accessibility relation R. It says that an information state s supports $\Diamond\phi$ iff for every world w in s, there is a non-empty substate t of accessible worlds from w such that ϕ is supported in t. The state s anti-supports $\Diamond\phi$ iff for every world w in s, the state consisting of accessible worlds from w anti-supports ϕ. The necessity modal operator $\Box$ can be defined as $\neg\Diamond\neg\phi$.

Aloni (Aloni (2022)) further proposed two state-based constraints on the accessibility relation R: indisputability and state-basedness. These constraints are used to capture deontic and epistemic modals respectively. Aloni defined the constraints as follow:

Definition 8 (State-based constraints)**:**

- *The accessibility relation R is indisputable in (M, s) iff* $\forall w, v \in s$*:* $R[w] = R[v]$
- *The accessibility relation R is state-based in (M, s) iff* $\forall w \in s$*:* $R[w] = s$

By this definition, it is straightforward to derive that if R is state-based, then it is also indisputable. On grounds of the constraints on relation R and Fact 1, some key results can be drawn from BSML.

Fact 2 (Some results in BSML)**:** [1]

- $\Box$*-free choice*

  $[\Box(p \vee q)]^+ \vDash \Diamond p \wedge \Diamond q$
- $\Diamond$*-free choice*

  $[\Diamond(p \vee q)]^+ \vDash \Diamond p \wedge \Diamond q$

1 In Aloni's system (BSML), free choice and ignorance inferences are taken to be semantics while they are assumed to be triggered by pragmatic effects but not of the conversational implicature kind. Rather they follow from something else that speakers do in conversation, namely the neglect zero effects. This is because some recent experiments have shown that these inferences are robust and generally more faster to process, and easier to acquire than regular conversational implicature, and a Gricean approach cannot account for this (see Marty et al. (2023); Ramotowska et al. (2022).

- *Wide scope free choice*

  $[\Diamond p \vee \Diamond q]^{+} \vDash \Diamond p \wedge \Diamond q$ *(R is indisputable)*
- *Ignorance inference (labeled as modal disjunction in BSML)*

  $[p \vee q]^{+} \vDash \Diamond p \wedge \Diamond q$ *(R is state-based)*
- *Upward monotonicity*

  $p \vDash p \vee q$, $\Diamond p \vDash \Diamond(p \vee q)$

The predication of upward monotonicity is only for the formulas without pragmatic enrichment. As shown in Anttila (2021), the NE-free fragment of BSML is equivalent to classical modal logic. So the monotonicity in BSML is preserved as a semantic property, implemented in the classical fragment of BSML. As illustrated by Figure 2.2, classical formulas and their pragmatically enriched variants are very different in BSML and give rise to different inferences. So free choice and ignorance effects cannot be derived by the NE-free fragment: $\Diamond(p \vee q) \nvDash \Diamond p \wedge \Diamond q$. While monotonicty will not be valid if there is a pragmatic intrusion, e.g. we can find counterexamples to show $\Diamond p \nvDash \Diamond[p \vee q]^{+}$, or $[\Diamond p]^{+} \nvDash \Diamond[p \vee q]^{+}$, where p is supported by an information state s but q is not.

Aloni further distinguished strong and weak notions of tautologies and contradictions.

	Weak	Strong
Tautologies	supported by every non-empty states	supported by every state
Contradictions	supported **only** by the empty state	never supported

Table 2.1 Tautologies and contradictions in BSML

For example, the atom NE is a weak tautology. The strong tautology is supported by every state, including the empty state. The classical tautology $p \vee \neg p$ is a strong tautology. While both $\neg$NE and $p \wedge \neg p$ are weak contradictions. A strong contradiction is never supported and it requires an empty and non-empty supporting state. So an example of strong contradiction would be NE $\wedge \neg$NE or NE $\wedge (p \wedge \neg p)$. In what follows, we use different notations to explicitly distinguish the weak and strong notions. We use $\top_w$ and $\bot_w$ for weak tautologies and contradictions respectively. And $\top_s$ and $\bot_s$ for strong tautologies and contradictions.

We have sketched the idea of BSML, in which pragmatic enrichment yields pragmatic effects of free choice and its variants. In the next section, we discuss the modal operator of epistemic *might* in BSML.

2.4.2 Epistemic possibilities

The discussed constraints on the accessibility relation give a team-based characterization of the properties of the relation. In particular, if we put the state-basedness constraint on R in (M, s), then R is in fact an equivalence relation (reflexive, transitive and symmetric) similar to the relation of classical **S5** system which is usually used to reason about knowledge.

Corollary 1: *The accessibility relation R is state-based in (M, s) only if R is an equivalence relation.*

As shown in Aloni (2022), the constraint of state-basedness on R provides an analysis of epistemic possibilities that simultaneously captures epistemic contradictions and non-factivity. This is not a trivial result, since it is *prima facie* difficult to reconcile the two principles in classical logic, as mentioned above.

Fact 3:

- *Epistemic contradiction*
  $\neg p \wedge \Diamond p \vDash \bot_w$ *(R is state-based)*
- *Non-factivity of epistemic possibility*
  $\Diamond p \nvDash p$ *(R is state-based)*

In BSML, epistemic contradictions are weak contradictions, which are only supported by empty states. Because an empty state would support both p and $\Diamond p$.

In this article, we propose a different analysis of epistemic possibility implemented in BSML, which also gives correct predictions with respect to epistemic contradictions and non-factivity. In addition, our analysis is more suitable to capture the interaction of epistemic *might* with knowledge and belief.

Inspired by the analysis of epistemic *might* in Veltman (1996), we propose that a statement *might* ϕ is acceptable if and only if ϕ is consistent with the information state of the relevant agent. An information state, as noted in Veltman et al. (1996), can be seen as a set of possibilities. Every non-empty subset of an information state provides certain possibilities with the potential to be knowledge. And a proposition supported by all possibilities in the information state is regarded as an epistemic necessity. Saying ϕ is consistent with the information state is to say that at least one non-empty subset of the information state supports that ϕ. Otherwise, it is not *might* ϕ. Therefore, we propose the following semantics of epistemic *might* which is denoted by $\blacklozenge$.

Definition 9 (Semantics of epistemic possibilities)**:**

$M, s \vDash \blacklozenge\phi \quad \textit{iff} \quad \exists s' \subseteq s \textit{ and } s' \neq \varnothing \textit{ s.t. } M, s' \vDash \phi$

$M, s \dashv \blacklozenge\phi \quad \textit{iff} \quad M, s \dashv \phi$

The formula $\blacklozenge\phi$ is supported by an information state s iff there is a non-empty substate of s supporting ϕ. And $\blacklozenge\phi$ is anti-supported by s iff ϕ is anti-supported by s, which means there is no possibility of ϕ in s. The epistemic necessity $\blacksquare$ is defined as $\neg\blacklozenge\neg\phi$.

Fact 4 (Epistemic contradiction with $\blacklozenge$)**:**

$\neg p \wedge \blacklozenge p \vDash \bot_s$

Proof: Suppose $M, s \vDash \neg p \wedge \blacklozenge p$. Then $M, s \vDash \neg p$ and $M, s \vDash \blacklozenge p$. The former implies $M, s \dashv p$, hence for every $w \in s$ $V(w, p) = 0$; and the latter implies that there is a non-empty subset s': $M, s' \vDash p$, so there exists $w \in s$, $V(w, p) = 1$. Contradiction! □

Notice that an epistemic contradiction with $\blacklozenge$ is a strong contradiction, i.e., not even the empty state supports it. This is different from epistemic contradictions with $\Diamond$ in BSML which are weak contradictions. This is because the assertability of $\blacklozenge$-formulas requires the information state s to be non-empty while the assertability of $\Diamond$ in BSML does not come with this requirement.

Fact 5 (Non-factivity of $\blacklozenge$)**:**

$\blacklozenge p \nvDash p$

Proof: We can construct a counterexample to falsify $\blacklozenge p \vDash p$. Suppose $s = \{w_1, w_2\}$ and $V(w_1, p) = 1$, $V(w_2, p) = 0$. So we have $M, s \vDash \blacklozenge p$ since $\{w_1\} \vDash p$. But $M, s \nvDash p$, as $V(w_2, p) = 0$. □

In our proposal, the analysis of *might* is non-relational. The accessibility relation does not play any role in the interpretation of epistemic possibilities. *Might* operates directly on the local information state. So we don't have to use the state-based constraint on R. This move will give us a straightforward account of the meaning of *might* under attitude verbs where *might* should be interpreted with respect to the information state of the subject of the attitude rather than the speaker.

In a Kripke-style model, we always interpret knowledge in terms of informational indistinguishability between possible worlds. To define a relational *might* in the scope of *know* or *believe*, we would need a state-based relation R with respect to the local information state (of the subject) rather than the global state (of the speaker). It is not totally trivial how to arrive at this notion in BSML. We would need to define different versions of the state-based constraint for different agents. Whereas this locality effect is straightforward in our approach given Definition 9 of $\blacklozenge$.

2.4.3 A sketch of the solution for Puzzle 1

As mentioned above, the NE-free fragment of BSML is equivalent to classical modal logic, but the free choice and ignorance principles can be applied only to pragmatically enriched formulas. We will call semantically valid the inferences validated by the NE-free fragment of BSML and pragmatically valid the inferences holding for pragmatically enriched formulas. In the next section we will extend BSML with a notion of knowledge which will give us a full account of Puzzle 1, repeated here as (25) and (26). We present here the core of the solution abstracting from the specifics of the interpretation of *know*. The monotonic inference from (25-a) to (25-b) ((26-b) to (26-c)) will be semantically valid. The ignorance (free choice) inference from (25-b) to (25-c) ((26-c) to (26-d)) instead will only be pragmatically valid. (25-c) ((26-d)) can only be derived by a pragmatically enriched version of (25-b) ((26-c)) which is not licensed by (25-a) ((26-b)).

(25) Version 1

a. Daiyu knows that the Netherlands is in Europe. $\texttt{KNOW}\ p$

b. Daiyu knows that the Netherlands is in Europe or Asia. $\texttt{KNOW}\ (p \vee \neg p)$

c. Daiyu knows that the Netherlands might be in Asia. $\texttt{KNOW}\ \blacklozenge\neg p$

d. The Netherlands might be in Asia. $\blacklozenge\neg p$

e. The Netherlands is in Europe. p

(26) Version 2

a. The Netherlands is in Europe. p

b. Daiyu knows that the Netherlands might be in Europe. $\texttt{KNOW}\ \blacklozenge p$

c. Daiyu knows that the Netherlands might be in Europe or Asia. $\texttt{KNOW}$ $\blacklozenge(p \vee \neg p)$

d. Daiyu knows that the Netherlands might be in Asia. $\texttt{KNOW}\ \blacklozenge\neg p$

e. The Netherlands might be in Asia. $\blacklozenge\neg p$

In this way, the ignorance principle, applied to $\texttt{KNOW}[(p \vee \neg p)]^+$, gives rise to (25-c). But the pragmatic enrichment of (25-b) cannot be derived from the assumption (25-a), so it is not assertable. It is same for the version 2. Therefore the puzzle can be resolved as

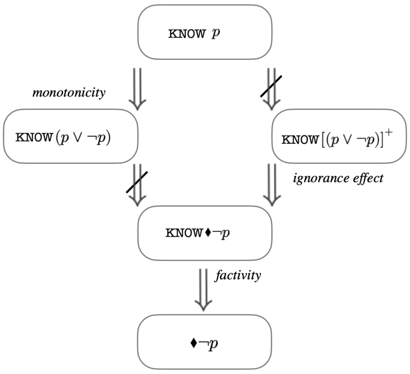


Figure 2.3 Strategy to Puzzle 1: version 1

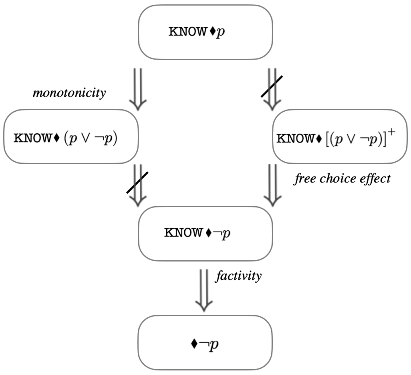


Figure 2.4 Strategy to Puzzle 1: version 2

a case of equivocation, and we can safely use the monotonicity and free choice principle. Hence, the WEM in the context of knowledge can be seen as an inference adhering to semantic principles while triggering certain pragmatic consequences.

In the next section, we propose an account of knowledge and belief in an extension of BSML using plausibility models. This account will highlight subtle distinctions between these two notions and also give us a strategy to solve Puzzle 2.

2.5 Belief and knowledge

2.5.1 Epistemic plausibility model for belief

Plausibility models are standard formal representations used in the literature on modeling belief revision, see for examples Spohn (1988); Stalnaker (2006). In Van Benthem (2007); Baltag and Smets (2016), an epistemic plausibility model was proposed to capture the dynamics of belief change and its interaction with knowledge. The idea underlying the model is that a *plausibility relation* is defined in the model to capture the agent's beliefs. If an agent was given the information of two worlds w and w', then she would believe the most plausible of the two. For example, if w is more plausible then the agent will believe w to be the actual world. Instead, if the two worlds are equivalently plausible then the agent will be indifferent between the worlds, namely the two worlds are doxastically indistinguishable for the agent.

The distinction between knowledge and belief can be studied from a dynamic perspective. Knowledge, as Van Benthem (2007) noted, is represented by describing information ranges for agents. The propositions containing agent's knowledge which are considered to be true will change only after the update of *hard information* (e.g. the public announcement of a fact). Whereas belief, compared with knowledge, is more volatile. It will change not only after hard information updating, but also after so-called *soft information* (e.g. hints of facial expressions) updating. In epistemic plausibility models, hard information updating induces possible worlds eliminations. While the effects of soft information updating are produced by changing the *plausibility relation* between worlds. Accordingly, the notion of belief is semantically weaker than knowledge, in the sense that the evaluation of propositions of belief only need to consider the most plausible alternatives.

We will use the epistemic plausibility model defined in Van Benthem (2007). For the sake of simplicity, we introduce the single agent version (see Section 7 for a discussion of the multi-agent extension). A standard **S5** epistemic model is a triple $\mathrm{M} = \langle W, \sim, V\rangle$. The relation $\sim$ is usually assumed to be an equivalence relation, and thus to give rise to equivalence classes $[w]$ for each $w \in W$. An epistemic plausibility model $\mathrm{M_p}$ is an extension of M obtained by adding a plausibility relation $\leq_w$ for every world in W. Intuitively, for every world v and v', $v \leq_w v'$ says that at the world w, v is at least as plausible as v'. Then the semantics of belief can be defined as follows.

- $\mathrm{M_p}, w \vDash [B]\phi$ iff $\mathrm{M_p}, v \vDash \phi$ for every $v \in Min_{\leq_w}([w])$
  where $Min_{\leq_w}([w]) := \{v \in [w] \mid \forall w' \in [w] : w' \leq_w v \Rightarrow v \leq_w w'\}$

The truth of $[B]\phi$ in w means ϕ is true in the $\leq_w$-minimal worlds of $[w]$, which is the most plausible part of $[w]$.

We will employ plausibility models to formalise belief in an extension of BSML. Since BSML models assertability conditions, this move will allow us to implement the idea that belief is weak also with respect to assertions, not only with respect to knowledge (Hawthorne et al. (2016)). Because in the model what we believe are propositions holding in the “best” or “most plausible” worlds which are epistemically accessible to us, whereas the assertion of ϕ requires ϕ to be supported by the whole information state.

2.5.2 The plausibility relation in BSEL

We extend the standard Kripke model, defined in Section 2.4, to an epistemic framework $\mathcal{M}$ by adding a plausibility ordering $\leq_w$. We call this model the Bilateral State-based Epistemic Logic (BSEL) model. The model $\mathcal{M}$ is now a tuple $\langle W, R, \leq_w, V\rangle$. The relation R is an equivalence relation, and we can also represent it by $\sim$. The equivalence class and plausibility ordering are defined similar to the epistemic plausibility model. Then the operator $\boldsymbol{[B]}$ can be defined for belief in $\mathcal{M}$.

Definition 10 (Belief)**:**

$M, s \vDash [\boldsymbol{B}]\phi$ *iff* $\forall w \in s, [w]_{dox} \vDash \phi$

$M, s \dashv [\boldsymbol{B}]\phi$ *iff* $\forall w \in s$, *there is a non-empty substate* $t \subseteq [w]_{dox} : t \dashv \phi$

where $[w]_{dox} = Min_{\leq_w}([w]) := \{v \in [w] \mid \forall w' \in [w] : w' \leq_w v \Rightarrow v \leq_w w'\}$, *and* $[w] = R[w]$.

The set $[w]_{dox}$ for every $w \in s$ is a minimal set of $[w]$, and $\bigcup_{w \in s} [w]_{dox}$ represents the doxastic state of an agent. This says that $[\boldsymbol{B}]\phi$ is supported by s iff ϕ is supported by the state consisting of the most plausible worlds in $[w]$ for every world $w \in s$. So doxastic formulas are interpreted with respect to doxastic state instead of information states s. The doxastic state in fact is generated from the information state by the accessibility relation and plausibility relation. Definition 10 gives us correct predictions concerning epistemic contradictions under belief.

Fact 6 (Epistemic contradictions under belief)**:**

i. $[\boldsymbol{B}](p \wedge \blacklozenge\neg p) \vDash \bot_w$

ii. $[\boldsymbol{B}]p \wedge [\boldsymbol{B}]\blacklozenge\neg p \vDash \bot_w$

Proof: i. Suppose $\mathcal{M}, s \vDash [\boldsymbol{B}](p \wedge \blacklozenge\neg p)$, which implies that $[w]_{dox} \vDash p \wedge \blacklozenge\neg p$ for every $w \in s$. By Fact 4, however, $p \wedge \blacklozenge\neg p \vDash \bot_s$. Strong contradictions cannot be supported by any state. So if $[\boldsymbol{B}](p \wedge \blacklozenge\neg p)$ is supported by s then s must be empty. Therefore we derive a weak contradiction.

ii. Suppose $\mathcal{M}, s \vDash [\boldsymbol{B}]p \wedge [\boldsymbol{B}]\blacklozenge\neg p$. This means that $\mathcal{M}, s \vDash [\boldsymbol{B}]p$ and $\mathcal{M}, s \vDash [\boldsymbol{B}]\blacklozenge\neg p$. So $[w]_{dox} \vDash p$, and $[w]_{dox} \vDash \blacklozenge\neg p$ for every $w \in s$. Take an arbitrary $w \in s$. The former means that for every $v \in [w]_{dox}$, $V(v, p) = 1$. Whereas the latter implies that there is a non-empty substate of $[w]_{dox}$ that supports $\neg p$, so there is at least one world $v \in [w]_{dox}$ such that $V(v, p) = 0$. This is impossible. Therefore, only an empty state can satisfy both $[\boldsymbol{B}]p$ and $[\boldsymbol{B}]\blacklozenge\neg p$. So a weak contradiction is derived.

□

It is worth noting that epistemic contradictions with $\blacklozenge$ are strong contradictions, but they become weak contradictions when embedded under belief. So in BSEL, the belief of an epistemic contradiction is weaker than its assertion (see Hawthorne et al. (2016)).

In this analysis, the locality of $\blacklozenge$ plays a role in the interpretation of $\blacklozenge$-formulas in the scope of $[\boldsymbol{B}]$. Those $\blacklozenge$-formulas are interpreted with respect to the local doxastic state rather than the global information state s. So when there is more than one agent, suppose a speaker and a subject of the aimed belief attribution, this setting allows us to express the possibilities believed by the subject but not necessarily entertained by the speaker. This is because the formula $[\boldsymbol{B}]\blacklozenge\phi$ does not require $\blacklozenge\phi$ to be supported by the global state of the speaker[1].

So far, we have defined the doxastic modality in BSEL employing the analyses of epistemic plausibility model for belief. Can we also define knowledge in BSEL based on it, like the classical **S**5 modality? Not in a straightforward way. This is because of a tension between the factivity of knowledge and the locality of the interpretation of *might*. We will take a closer look at this issue in the following section.

2.5.3 A problem for factivity

In a BSEL model $\mathcal{M}$, the accessibility relation $\boldsymbol{R}$ is assumed to be an equivalence relation, i.e. $\boldsymbol{R}$ is reflexive, transitive and symmetric. Consider now the **S**5 modality $[\sim]$ defined as follows: s supports $[\sim]\phi$ iff ϕ is supported by $[w]$ for every $w \in s$. Can we take this modality to represent knowledge in our framework as it is usually done in plausibility model accounts? This seems *prima facie* a reasonable option. Because of the reflexivity of $\boldsymbol{R}$, it must be that $w \in [w]$, which provides a traditional way to capture factivity of knowledge. Reflexivity alone however is not enough to account for the factivity in our framework. Consider the following counterexample.

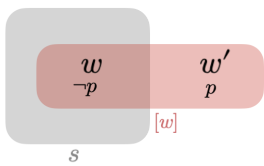


Figure 2.5 A counterexample for factivity

In Figure 2.5, we depict in light grey the information state s and in red the state $[w]$. In the model, there is a single world w in s, and a world w' which is accessible from w but is not part of s. Given our definition of $\blacklozenge$ it follows that $\mathcal{M}, s \vDash [\sim]\blacklozenge p$, but $s \nvDash \blacklozenge p$. This is because, for the interpretation of $[\sim]\blacklozenge p$, $\blacklozenge p$ in the scope of $[\sim]$ is interpreted with respect to $[w]$, in which there is a substate $\{w'\}$ supporting p. Whereas, unembedded $\blacklozenge p$ is interpreted with respect to s where p is anti-supported. Hence, we lose factivity of

1 In principle, we cannot express multi-agent scenarios in the current analysis. We distinguish the speaker and the subject in a way that we assume the information state that is the current evaluation point represents the speaker's information, while the generated states (by accessibility relation) represent the subject's information. This allows us to distinguish at least two agents. When expressing the speaker's own knowledge and beliefs, we need to assume the accessibility relation to be state-based. More discussion on multi-agent can be found in Section 2.8 and Yan et al. (2023).

knowledge for epistemic possibilities: $[\sim]\blacklozenge p \nvDash \blacklozenge p$.

This problem motivates us to propose a new interpretation of knowledge which will capture factivity also for *know-might* sentences. As for the **S**5 modality, we will call it an **informational modality**[1] and denote it by $[\boldsymbol{I}]$. The formula $[\boldsymbol{I}]\phi$ means that every piece of information the agent has supports ϕ. As we will see, this notion will be particularly useful in the multi-agent setting, in which we need to represent the informational relationship between information states of different individuals.

Definition 11 (Informational modality)**:**

$\mathcal{M}, s \vDash [\boldsymbol{I}]\phi$ *iff* $\forall w \in s, [w] \vDash \phi$

$\mathcal{M}, s \dashv [\boldsymbol{I}]\phi$ *iff* $\forall w \in s$: *there is a non-empty subset* $t \subseteq [w]$ *and* $t \dashv \phi$

This says that s supports $[\boldsymbol{I}]\phi$ if and only if for every $w \in s$ the equivalence class $[w]$ supports ϕ. And s anti-supports $[\boldsymbol{I}]\phi$ if and only if there is a piece information the agent has which anti-supports ϕ.

The modality $[\boldsymbol{I}]$ does not represent knowledge, in contrast to what logicians assume in standard epistemic logic. There are two reasons for this move, both connected to factivity. One is to avoid the problem explained above and capture the factivity of knowledge also for *know-might* sentences. The second motivation concerns the projection of the factivity inference under logical operations, in particular under negation. For example, the following inference has been observed by linguists to be correct but it is not captured if we analyse knowledge by the plain **S**5 modality.

(27) Daiyu does not know that Baoyu is in Rongguofu. $\rightsquigarrow$ Baoyu is in Rongguofu.

To capture these facts, in the following section we propose to treat factivity as a presupposition.

2.5.4 Factivity as presupposition

As Ludlow (2005) noted, the notion of knowledge is best reflected in the way we commonly talk about knowledge. So factivity can be defended by appealing to the uses of the verb *know* in language. Let us first see how the factivity of knowledge is characterized from a linguistic perspective.

Linguists normally assume that *know* triggers the presupposition of the truth of any

1 We treat this notion as informational knowledge or declarative knowledge, which is from the information science. This notion indicates a description of pieces of information generated from data.

declarative complement it may take (e.g.Spector and Egré (2015)). The sentence "α knows S" presupposes, rather than asserts, that S is true. This implies, for example, that the sentence "Daiyu knows it is raining." has no truth value in those worlds where it is not raining (instead of being false in those worlds). Another defining characteristics of presupposition concerns their projective behaviour. While plain inferences are canceled under negation, presuppositions normally project. As mentioned, the factivity inference under knowledge seems to follow this pattern, for example "Daiyu doesn't know it is raining." implies "It is raining". As an illustration consider Spector and Egré (2015) entry for knowledge:

$$[\![know]\!]^{w} = \lambda\phi.\lambda\alpha : \phi(w) = 1.\ \alpha \text{ believes } \phi \text{ in } w.$$

In this semantics, the verb *know* indicates that its complement is presupposed to be true (expressed by $\phi(w) = 1$). and the agent believes this complement. This account implies that *know*-sentences and their negations have a truth value only in a situation where its complement is true. In our state-based framework, we formulate assertability instead of truth. Thus, we propose that *know*-sentences can be asserted only in situations where also their complement is assertable. As in semantics of Spector and Egré (2015), we will therefore model the factivity of knowledge as a presupposition. But, in our definition we will get rid of the belief condition which seems too weak to guarantee knowledge, as we argued above. Instead, we will replace it with the condition of informational modality.

To implement these ideas, we extend our language a new complex formula ϕ_ψ which should be read as ϕ presupposes ψ. The semantics is given in the following definition.

Definition 12 (Presupposition)**:**

$\mathcal{M}, s \vDash \phi_\psi$ *iff* $s \vDash \phi$ *and* $s \vDash \psi$

$\mathcal{M}, s \mathrel{=\!\!|} \phi_\psi$ *iff* $s \mathrel{=\!\!|} \phi$ *and* $s \vDash \psi$

Notice that, the rejectability condition predicts that presuppositions are preserved under negation[1].

$$\neg\phi_\psi \vDash \psi$$

We can now define knowledge in terms of presupposition and informational modality.

Definition 13 (Knowledge)**:**

$$[K]\phi := ([I]\phi)_\phi$$

Based on this definition, we can derive the factivity of knowledge for factual complements but also for epistemic possibilities $\blacklozenge\phi$.

1 Presuppositions have been argued to project under more operators (see Karttunen (1973)). Our analysis will not yet account for this. In future work we will consider dynamic versions of BSML where the full projective behaviour of presuppositions can be accounted for (see Aloni (2023)).

Fact 7 (Factivity of knowledge for epistemic possibilities)**:**

$$[\boldsymbol{K}]\blacklozenge p \vDash \blacklozenge p$$

Proof: Suppose a model $\mathcal{M}, s \vDash [\boldsymbol{K}]\blacklozenge p$, which means $s \vDash ([\boldsymbol{I}]\blacklozenge p)_{\blacklozenge p}$. It implies that $s \vDash \blacklozenge p$ and $s \vDash [\boldsymbol{I}]\blacklozenge p$. Therefore, $\mathcal{M}, s \vDash \blacklozenge p$. □

In addition, we predict the projection of the factivity presupposition under negation, a pragmatic effect which cannot be derived by classical epistemic logic.

(28) a. Daiyu doesn't know that Baoyu is wearing glasses. $\rightsquigarrow$ Baoyu is wearing glasses.

b. Daiyu doesn't know that Baoyu might be in Beijing. $\rightsquigarrow$ Baoyu might be in Beijing.

Fact 8 (Factive inferences under negation)**:**

$\neg[\boldsymbol{K}]\phi \vDash \phi$

Proof: This can be derived directly by the definitions of knowledge and presupposition. □

In the next section, we put together all the ingredients discussed so far and present the full picture of our final proposal, a Bilateral State-based Epistemic Logic (BSEL).

2.6 Bilateral state-based epistemic logic (BSEL)

2.6.1 The framework

Our target language $\mathcal{L}_E$ is based on the propositional language $\mathcal{L}$ of BSML, enriched by the epistemic modalities $[\boldsymbol{B}]$, $[\boldsymbol{I}]$, $\blacklozenge$ and presupposition ϕ_ψ.

Definition 14 (Language $\mathcal{L}_E$ of BSEL)**:**

$$\phi ::= p \mid \neg\phi \mid \phi \wedge \phi \mid \phi \vee \phi \mid \text{NE} \mid [\boldsymbol{B}]\phi \mid [\boldsymbol{I}]\phi \mid \blacklozenge\phi \mid \phi_\phi$$

As introduced in Section 2.4.1, NE is the non-emptiness atom from team logic; $[\boldsymbol{B}]\phi$ is the doxastic modality, which expresses that the agent believes ϕ; $[\boldsymbol{I}]\phi$ stands for the informational modality of ϕ, i.e. every piece of information the agent has supports ϕ; $\blacklozenge\phi$ is the epistemic possibility of ϕ; and the final one ϕ_ϕ indicates that ϕ presupposes ϕ. The modality of knowledge in $\mathcal{L}_E$ is defined as a derived notion: $[\boldsymbol{K}]\phi := ([\boldsymbol{I}]\phi)_\phi$. $[\boldsymbol{K}]\phi$ means that ϕ is presupposed to be assertable and it is supported in every piece of information the agent has.

Definition 15 (Pragmatic enrichment based on $\mathcal{L}_E$)**:** *We define the pragmatic enrichment for the new formulas in BSEL.*

- $[\blacklozenge\phi]^+ = \blacklozenge[\phi]^+ \wedge \mathrm{NE}$
- $[[\boldsymbol{B}]\phi]^+ = [\boldsymbol{B}][\phi]^+ \wedge \mathrm{NE}$
- $[[\boldsymbol{I}]\phi]^+ = [\boldsymbol{I}][\phi]^+ \wedge \mathrm{NE}$
- $[\phi_\psi]^+ = [\phi]^+_{[\psi]^+} \wedge \mathrm{NE}$

The pragmatic enrichment of knowledge formulas can be defined in term of the pragmatic enrichment of $[[\boldsymbol{I}]\phi]^+$ and $[\phi_\psi]^+$: $[[\boldsymbol{K}]\phi]^+ = [([\boldsymbol{I}]\phi)_\phi]^+$

The epistemic model $\mathcal{M}$ for $\mathcal{L}_E$ is a tuple $\langle W, R, \leq_w, V\rangle$. There are two relations over the worlds, the equivalence relation R and the plausibility relation $\leq_w$. Based on these two relations, we have two kinds of sets of possible worlds, the equivalence class $[w]$ and the $\leq_w$-minimal set $[w]_{dox}$ for every $w \in s$. Accordingly, in an epistemic model, in addition to information state s formulas can be also interpreted relative to $[w]$ and $[w]_{dox}$. The following figure shows a model of the three different states.

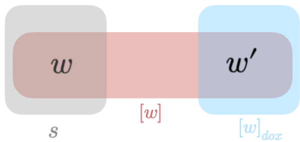


Figure 2.6 Three kinds of sets of possible worlds

In Figure 2.6, the light grey area stands for the information state s; the red area stands for the equivalence class of w and the blue area represents the most plausible worlds in $[w]$.

Definition 16 (Semantic clause for new modals)**:**

$\mathcal{M}, s \vDash \blacklozenge\phi$ *iff* $\exists s' \subseteq s$ *and* $s' \neq \varnothing$ *s.t.* $\mathcal{M}, s' \vDash \phi$
$\mathcal{M}, s \dashv \blacklozenge\phi$ *iff* $\mathcal{M}, s \dashv \phi$
$\mathcal{M}, s \vDash [\boldsymbol{I}]\phi$ *iff* $\forall w \in s : [w] \vDash \phi$
$\mathcal{M}, s \dashv [\boldsymbol{I}]\phi$ *iff* $\forall w \in s$*: there is a non-empty subset* $t \subseteq [w]$ *and* $t \dashv \phi$
$\mathcal{M}, s \vDash [\boldsymbol{B}]\phi$ *iff* $\forall w \in s$, $[w]_{dox} \vDash \phi$
$\mathcal{M}, s \dashv [\boldsymbol{B}]\phi$ *iff* $\forall w \in s$, *there is non-empty substate* $t \subseteq [w]_{dox} : t \dashv \phi$
$\mathcal{M}, s \vDash \phi_\psi$ *iff* $s \vDash \phi$ *and* $s \vDash \psi$
$\mathcal{M}, s \dashv \phi_\psi$ *iff* $s \dashv \phi$ *and* $s \vDash \psi$

The necessity $\blacksquare$ is defined as $\neg\blacklozenge\neg$. Different modals are interpreted with respect to different sets of possible worlds s, $[w]$ or $[w]_{dox}$. The interpretation of epistemic possibility $\blacklozenge$ is non-relational, so it can be interpreted relative to the all three kinds of sets. Informational modality is interpreted w.r.t the equivalence class $[w]$, and belief is interpreted w.r.t the equivalence class $[w]$ and the $\leq_w$-minimal set $[w]_{dox}$.

2.6.2 Solution to Puzzle 1

We have now a full solution to Puzzle 1, as illustrated in Figures 2.7 and 2.8. Consider again the examples (16) and (17). As noted, "Daiyu knows that the Netherlands might be in Asia." cannot be draw semantically from "Daiyu knows that the Netherlands is (might be) in Europe or Asia.", instead it follows from the pragmatically enriched version of the proposition. An illustration is given by the following figures. Let p be the proposition "the Netherlands is in Europe".

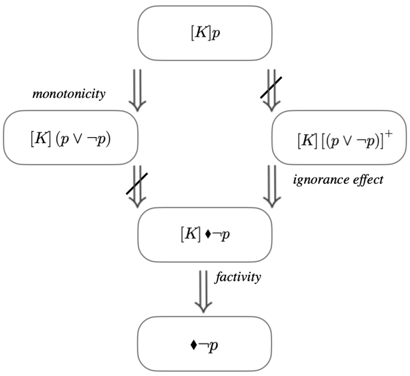


Figure 2.7 Puzzle 1: version 1

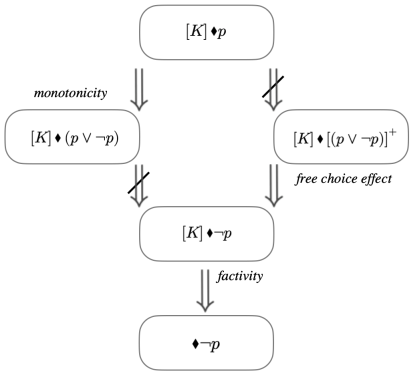


Figure 2.8 Puzzle 1: version 2

From the premises $[\boldsymbol{K}]p$ or $[\boldsymbol{K}]\blacklozenge p$ we cannot derive $[\boldsymbol{K}]\blacklozenge\neg p$ because the enriched version of $[\boldsymbol{K}]\ (p \vee \neg p)$ and $[\boldsymbol{K}]\blacklozenge(p \vee \neg p)$ are not assertable.

2.6.3 Solution to Puzzle 2

Let's go back to the epistemic entailment thesis (EET). In BSEL, we use plausibility ordering to model belief, and presupposition as well as informational modality to define knowledge. The idea is that belief quantifies over a smaller set of alternatives than knowledge, which guarantees the entailment from knowledge to belief for factual propositions but not for *might*-sentences. The following figure illustrates a counterexample.

In the model shown by Figure 2.9, the information state s has a single world w in which p holds. And there is a world w' accessible from w, so $w' \in [w]$. The world w' is the most plausible world, and $V(w', \neg p)$=1. By the model, we have $s \vDash [\boldsymbol{K}]\blacklozenge p$ since

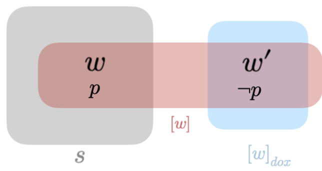


Figure 2.9 $[\boldsymbol{K}]\blacklozenge p \nvDash [\boldsymbol{B}]\blacklozenge p$

the non-empty substate $\{w\}$ of $[w]$, i.e. s itself, supports p. However $s \nvDash [\boldsymbol{B}]\blacklozenge p$, since $[w]_{dox} \dashv p$.

In this way, we have blocked the entailment from *know-might* to *believe-might* sentences which gives us a solution to Puzzle 2.

In addition, we further observe that the EET is blocked also for pragmatically enriched formulas in our framework, which can be illustrated by the following counterexample.

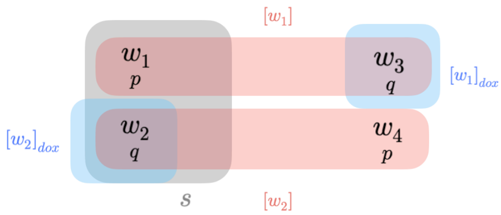


Figure 2.10 $[\boldsymbol{K}][(p \vee q)]^+ \nvDash [\boldsymbol{B}][(p \vee q)]^+$

In the model shown by Figure 2.10, the information state $s = \{w_1, w_2\}$, and $[w_1] = \{w_1, w_3\}$, $[w_2] = \{w_2, w_4\}$. The most plausible world in $[w_1]$ is w_3, and the most plausible world in $[w_2]$ is w_2. The proposition p holds in w_1 and w_4, and q holds in w_2 and w_3. By the definition of knowledge, it can be derived that $s \vDash [\boldsymbol{K}][(p \vee q)]^+$. Since $s \vDash [(p \vee q)]^+$ and $s \vDash [\boldsymbol{I}][(p \vee q)]^+$. However, neither $[w_1]_{dox}$ nor $[w_2]_{dox}$ supports p. So $s \nvDash [\boldsymbol{B}][(p \vee q)]^+$.

This observation can be evidenced by the following reasoning.

> *Suppose Daiyu knows that Baoyu is in Rongguofu or Ningguofu. So, pragmatically, Daiyu knows that Baoyu might be in Ningguofu. But Daiyu does not believe that Baoyu might be in Ningguofu. She only believes that Baoyu is in Rongguofu. Therefore she does not believe that Baoyu is in Rongguofu or Ningguofu.*

Therefore, we propose that the entailment from knowledge to belief holds but not in all cases. The epistemic entailment thesis holds only for the $\blacklozenge$–free and NE-free fragment of the language:

$$[\boldsymbol{K}]\phi \vDash [\boldsymbol{B}]\phi, \text{ if } \phi \text{ is a } \blacklozenge\text{-free and NE-free formula.}$$

2.7 More results

2.7.1 Non-classical properties of epistemic *might*

The framework BSEL we propose in this chapter makes similar predictions as BSML on facts about unembedded epistemic possibility. But as shown above unembedded epistemic contradictions are here strong contradictions rather than weak.

Result 1:

- *Epistemic contradiction*
  $\phi \wedge \blacklozenge\neg\phi \vDash \bot_s$
- *Non-factivity of epistemic possibility*
  $\blacklozenge\phi \nvDash \phi$

In addition, using pragmatic enrichment, we do not lose the advantage of BSML on predicting pragmatic effects. To derive these results, we need to extend Fact 1, showing the cases of the new formulas in BSEL.

Fact 9 (Extension of Fact 1)**:**

- $\mathcal{M}, s \vDash [\phi]^+ \Rightarrow \mathcal{M}, s \vDash \phi$
- $\mathcal{M}, s \mathrel{=\!\!|} [\phi]^+ \Rightarrow \mathcal{M}, s \mathrel{=\!\!|} \phi$

Proof: We only show the cases when ϕ has the form of $\blacklozenge\phi$, $[\boldsymbol{B}]\phi$, $[\boldsymbol{I}]\phi$ and ϕ_ψ.

By double induction on the complexity of ϕ

i. When $\phi = \blacklozenge\phi$
   - $\mathcal{M}, s \vDash [\blacklozenge\phi]^+$, which means that $\mathcal{M}, s \vDash \blacklozenge[\phi]^+$ and $s \neq \emptyset$. So there is non-empty substate $t \subseteq s$, $t \vDash [\phi]^+$. By induction hypothesis, $t \vDash \phi$. Therefore $\mathcal{M}, s \vDash \blacklozenge\phi$.
   - $\mathcal{M}, s \mathrel{=\!\!|} [\blacklozenge\phi]^+$, which means that $\mathcal{M}, s \mathrel{=\!\!|} \blacklozenge[\phi]^+$ and $s \neq \emptyset$. So $s \mathrel{=\!\!|} [\phi]^+$. By induction hypothesis, $s \mathrel{=\!\!|} \phi$. Therefore $\mathcal{M}, s \mathrel{=\!\!|} \blacklozenge\phi$.

ii. When $\phi = [\boldsymbol{B}]\phi$
   - $\mathcal{M}, s \vDash [[\boldsymbol{B}]\phi]^+$, which means that $\mathcal{M}, s \vDash [\boldsymbol{B}][\phi]^+$ and $s \neq \emptyset$. So for every $w \in s$, $[w]_{dox} \vDash [\phi]^+$. By induction hypothesis, for every $w \in s$, $[w]_{dox} \vDash \phi$. Therefore $\mathcal{M}, s \vDash [\boldsymbol{B}]\phi$.
   - $\mathcal{M}, s \mathrel{=\!\!|} [[\boldsymbol{B}]\phi]^+$, which means that $\mathcal{M}, s \mathrel{=\!\!|} [\boldsymbol{B}][\phi]^+$ and $s \neq \emptyset$. So for every $w \in s$, there is a non-empty substate $t \subseteq [w]_{dox} : t \mathrel{=\!\!|} [\phi]^+$. By induction

hypothesis, $t \dashv \phi$. Therefore $\mathcal{M}, s \dashv [\boldsymbol{B}]\phi$.

iii. When $\phi = [\boldsymbol{I}]\phi$

- $\mathcal{M}, s \vDash [[\boldsymbol{I}]\phi]^+$, which means that $\mathcal{M}, s \vDash [\boldsymbol{I}][\phi]^+$ and $s \neq \varnothing$. So for every $w \in s$, $[w] \vDash [\phi]^+$. By induction hypothesis, for every $w \in s$, $[w] \vDash \phi$. Therefore $\mathcal{M}, s \vDash [\boldsymbol{I}]\phi$.
- $\mathcal{M}, s \dashv [[\boldsymbol{I}]\phi]^+$, which means that $\mathcal{M}, s \dashv [\boldsymbol{I}][\phi]^+$ and $s \neq \varnothing$. So for every $w \in s$, there is a non-empty subset $t \subseteq [w]$ and $t \dashv [\phi]^+$. By induction hypothesis, $t \dashv \phi$. Therefore $\mathcal{M}, s \dashv [\boldsymbol{I}]\phi$.

iv. When $\phi = \phi_\psi$

- $\mathcal{M}, s \vDash [\phi_\psi]^+$, which means that $\mathcal{M}, s \vDash [\phi]^+_{[\psi]^+}$ and $s \neq \varnothing$. So $s \vDash [\phi]^+$ and $s \vDash [\psi]^+$. By induction hypothesis, $s \vDash \phi$ and $s \vDash \psi$. Therefore $\mathcal{M}, s \vDash \phi_\psi$.
- $\mathcal{M}, s \dashv [\phi_\psi]^+$, which means that $\mathcal{M}, s \dashv [\phi]^+_{[\psi]^+}$ and $s \neq \varnothing$. So $s \dashv [\phi]^+$ and $s \vDash [\psi]^+$. By induction hypothesis, $s \dashv \phi$ and $s \vDash \psi$. Therefore $\mathcal{M}, s \dashv \phi_\psi$.

□

Now we can derive epistemic (wide scope) free choice and ignorance inferences, but without the state-based constraint on accessibility relation $\boldsymbol{R}$.

Result 2:

- $\blacksquare$-*free choice*
  $[\blacksquare(\phi \vee \psi)]^+ \vDash \blacklozenge\phi \wedge \blacklozenge\psi$
- $\blacklozenge$-*free choice*
  $[\blacklozenge(\phi \vee \psi)]^+ \vDash \blacklozenge\phi \wedge \blacklozenge\psi$
- *Wide scope free choice*
  $[\blacklozenge\phi \vee \blacklozenge\psi]^+ \vDash \blacklozenge\phi \wedge \blacklozenge\psi$
- *Ignorance inference*
  $[\phi \vee \psi]^+ \vDash \blacklozenge\phi \wedge \blacklozenge\psi$

Proof:

- $\blacksquare$-*free choice*
  Suppose $\mathcal{M}, s \vDash [\blacksquare(\phi \vee \psi)]^+$. It implies that $\mathcal{M}, s \vDash \blacksquare[\phi \vee \psi]^+ \wedge \mathrm{NE}$, namely $\mathcal{M}, s \vDash \neg\blacklozenge\neg[\phi \vee \psi]^+$ and $s \neq \varnothing$. So $\mathcal{M}, s \dashv \blacklozenge\neg[\phi \vee \psi]^+$, which means $s \dashv \neg[\phi \vee \psi]^+$. This implies that $s \vDash [\phi \vee \psi]^+$. So $\mathcal{M}, s \vDash ([\phi]^+ \wedge \mathrm{NE}) \vee ([\psi]^+ \wedge \mathrm{NE})$. It follows that there are two non-empty substates $t_1, t_2 : t_1 \cup t_2 = s$, $t_1 \vDash [\phi]^+$ and $t_2 \vDash [\psi]^+$.

By Fact 9, $t_1 \vDash \phi$ and $t_2 \vDash \psi$. As a result $s \vDash \blacklozenge\phi \wedge \blacklozenge\psi$.

- *$\blacklozenge$-free choice*
  Suppose $\mathcal{M}, s \vDash [\blacklozenge(\phi \vee \psi)]^+$. It implies that $\mathcal{M}, s \vDash \blacklozenge[(\phi \vee \psi)]^+ \wedge \texttt{NE}$, namely $\mathcal{M}, s \vDash \blacklozenge[(\phi \vee \psi)]^+$ and $s \neq \varnothing$. This means that $s \vDash \blacklozenge(([\phi]^+ \wedge \texttt{NE}) \vee ([\psi]^+ \wedge \texttt{NE}))$, which implies that there is a non-empty substate $s' \subseteq s$, $s' \vDash ([\phi]^+ \wedge \texttt{NE}) \vee ([\psi]^+ \wedge \texttt{NE})$. So there are two non-empty substates t_1 and t_2, $t_1 \cup t_2 = s'$ and $t_1 \vDash [\phi]^+$ and $t_2 \vDash [\psi]^+$. By Fact 9, $t_1 \vDash \phi$ and $t_2 \vDash \psi$. Since $t_1, t_2 \subseteq s' \subseteq s$, we derive that $s \vDash \blacklozenge\phi$ and $s \vDash \blacklozenge\psi$.
- *Wide scope free choice*
  Suppose $\mathcal{M}, s \vDash [\blacklozenge\phi \vee \blacklozenge\psi]^+$, which means that $\mathcal{M}, s \vDash ([\blacklozenge\phi]^+ \wedge \texttt{NE}) \vee ([\blacklozenge\psi]^+ \wedge \texttt{NE})$. So there are two non-empty substate t_1, t_2: $t_1 \cup t_2 = s$, $t_1 \vDash [\blacklozenge\phi]^+$ and $t_2 \vDash [\blacklozenge\psi]^+$. By Fact 9, it follows that $t_1 \vDash \blacklozenge\phi$ and $t_2 \vDash \blacklozenge\psi$. So $s \vDash \blacklozenge\phi$ and $s \vDash \blacklozenge\psi$.
- *Ignorance inference*
  Suppose $\mathcal{M}, s \vDash [\phi \vee \psi]^+$, which means that $\mathcal{M}, s \vDash ([\phi]^+ \wedge \texttt{NE}) \vee ([\psi]^+ \wedge \texttt{NE})$. So there are two non-empty substate t_1, t_2: $t_1 \cup t_2 = s$, $t_1 \vDash [\phi]^+$ and $t_2 \vDash [\psi]^+$. By Fact 9, it follows that $t_1 \vDash \phi$ and $t_2 \vDash \psi$. So $s \vDash \blacklozenge\phi$ and $s \vDash \blacklozenge\psi$.

$\square$

In fact, by the semantics of $\blacklozenge$, if a pragmatically enriched disjunction can be verified, then we can always derive a conjunction of each disjunct in the scope of an epistemic possibility.

In Holliday and Mandelkern (2022), the authors observe the contrast between the following sentences:

(29) a. Sue might be the winner and she might not be, and either she is the winner or she isn't.

b. # Sue might not be the winner and she is the winner, or else Sue might be the winner and she isn't the winner.

Then sentence in (29-a) expresses an ignorance, but (29-b) is argued to be infelicitous. However, in the event that distributivity holds true, namely the logical equivalence between $\phi \wedge (\psi \vee \theta)$ and $(\phi \wedge \psi) \vee (\phi \wedge \theta)$, it follows that reference (29-a) implies reference (29-b).

We can account for the contrast and capture the infelicity of (29-b) by $\blacklozenge$[1].

(30) a. $(\blacklozenge p \wedge \blacklozenge\neg p) \wedge (p \vee \neg p)$

b. # $(\blacklozenge\neg p \wedge p) \vee (\blacklozenge p \wedge \neg p)$

While the formula in (30-a) is assertable by the definition of $\blacklozenge$, the formula in (30-b) is not. This is due to the fact that the substate supporting p will not support $\neg p$, and consequently, will not support $\blacklozenge\neg p$.

2.7.2 Interaction between epistemic *might* and other attitude verbs

As discussed in Section 2.2, epistemic free choice effects are still generated under *know* and *believe*. The semantics of belief, informational modality and knowledge in BSEL gives correct predictions of these phenomena.

Result 3:

- *Free choice under attitude verbs*
  $\boldsymbol{[I]}[\blacklozenge(\phi \vee \psi)]^+ \vDash \boldsymbol{[I]}\blacklozenge\phi \wedge \boldsymbol{[I]}\blacklozenge\psi$
  $\boldsymbol{[B]}[\blacklozenge(\phi \vee \psi)]^+ \vDash \boldsymbol{[B]}\blacklozenge\phi \wedge \boldsymbol{[B]}\blacklozenge\psi$
  $\boldsymbol{[K]}[\blacklozenge(\phi \vee \psi)]^+ \vDash \boldsymbol{[K]}\blacklozenge\phi \wedge \boldsymbol{[K]}\blacklozenge\psi$

Proof:

- Free choice under **[*I*]**
  Suppose $\mathcal{M}, s \vDash \boldsymbol{[I]}[\blacklozenge(\phi \vee \psi)]^+$, which means that $\forall w \in s : [w] \vDash [\blacklozenge(\phi \vee \psi)]^+$. This implies that $[w] \neq \varnothing$ and $\exists [w]' \subseteq [w]$ and $[w]' \neq \varnothing$: $[w]' \vDash ([\phi]^+ \wedge \mathrm{NE}) \vee ([\psi]^+ \wedge \mathrm{NE})$. So there are non-empty $t, t' : t \cup t' = [w]'$, $t \vDash [\phi]^+$ and $t' \vDash [\psi]^+$. By Fact 9, $t \vDash \phi$ and $t' \vDash \psi$. Since $t, t' \subseteq [w]$, $[w] \vDash \blacklozenge\phi$ and $[w] \vDash \blacklozenge\psi$. So $\mathcal{M}, s \vDash \boldsymbol{[I]}\blacklozenge\phi$ and $\mathcal{M}, s \vDash \boldsymbol{[I]}\blacklozenge\psi$.
- Free choice under **[*B*]**
  Suppose $\mathcal{M}, s \vDash \boldsymbol{[B]}[\blacklozenge(\phi \vee \psi)]^+$, which means that $\forall w \in s : [w]_{dox} \vDash [\blacklozenge(\phi \vee \psi)]^+$. Then similarly to the last proof we can show that $\forall w \in s : [w]_{dox} \vDash \blacklozenge\phi$ and $[w]_{dox} \vDash \blacklozenge\psi$. Therefore, $\mathcal{M}, s \vDash \boldsymbol{[B]}\blacklozenge\phi$ and $\mathcal{M}, s \vDash \boldsymbol{[B]}\blacklozenge\psi$.
- Free choice under **[*K*]**
  Suppose $\mathcal{M}, s \vDash \boldsymbol{[K]}[\blacklozenge(\phi \vee \psi)]^+$, which means $\mathcal{M}, s \vDash (\boldsymbol{[I]}[\blacklozenge(\phi \vee \psi)]^+)_{[\blacklozenge(\phi \vee \psi)]^+}$. This implies that $s \vDash \boldsymbol{[I]}[\blacklozenge(\phi \vee \psi)]^+$ and $s \vDash [\blacklozenge(\phi \vee \psi)]^+$. As shown in the first

1 We appreciate Yichi Zhang for bringing the data to us.

proof, the former implies that $s \vDash \mathbf{[}\boldsymbol{I}\mathbf{]}\blacklozenge\phi$ and $s \vDash \mathbf{[}\boldsymbol{I}\mathbf{]}\blacklozenge\psi$. And by the fact of $\blacklozenge$-free choice, the latter implies that $s \vDash \blacklozenge\phi$ and $s \vDash \blacklozenge\psi$. Therefore, $\mathcal{M}, s \vDash \mathbf{[}\boldsymbol{I}\mathbf{]}\blacklozenge\phi$ and $\mathcal{M}, s \vDash \blacklozenge\phi$, which means that $\mathcal{M}, s \vDash \mathbf{[}\boldsymbol{K}\mathbf{]}\blacklozenge\phi$. In the same way, we can prove $\mathcal{M}, s \vDash \mathbf{[}\boldsymbol{K}\mathbf{]}\blacklozenge\psi$.

□

It is worth noting that, we can define the following form of duality for the operator $\mathbf{[}\boldsymbol{I}\mathbf{]}$ and $\mathbf{[}\boldsymbol{B}\mathbf{]}$.

Result 4:

- $\mathbf{[}\boldsymbol{I}\mathbf{]}\blacklozenge\phi \Leftrightarrow \langle \boldsymbol{I} \rangle\phi$, where $\langle \boldsymbol{I} \rangle\phi := \neg[\boldsymbol{I}]\neg\phi$
- $\mathbf{[}\boldsymbol{B}\mathbf{]}\blacklozenge\phi \Leftrightarrow \langle \boldsymbol{B} \rangle\phi$, where $\langle \boldsymbol{B} \rangle\phi := \neg[\boldsymbol{B}]\neg\phi$

This shows that embedded *might*-sentences could be expressed in the original BSML as diamond versions of $\Box$. We would lose however a uniform account of *might* which only in the present system can be uniformly translated as $\blacklozenge$.

To give a solution to Puzzle 2, the entailment from knowledge to belief is blocked by $\blacklozenge$-formulas. Besides we observe that it is blocked by pragmatically enriched formulas. In addition, we can capture the infelicity of epistemic contradictions in the scope of *believe*.

Result 5:

- *Entailment from knowledge to belief*
  $[\boldsymbol{K}]\phi \vDash [\boldsymbol{B}]\phi$, if ϕ is $\blacklozenge$-free and NE-free
  But $[\boldsymbol{K}]\blacklozenge\phi \nvDash [\boldsymbol{B}]\blacklozenge\phi$, $[\boldsymbol{K}][\phi]^+ \nvDash [\boldsymbol{B}][\phi]^+$
- *Believing an epistemic contradiction*
  $[\boldsymbol{B}](\phi \wedge \blacklozenge\neg\phi) \vDash \bot_w$
  $[\boldsymbol{B}]\phi \wedge [\boldsymbol{B}]\blacklozenge\neg\phi \vDash \bot_w$
  $[\boldsymbol{B}]\phi \wedge [\boldsymbol{K}]\blacklozenge\neg\phi \nvDash \bot$

We predict that $[\boldsymbol{B}]\phi \wedge [\boldsymbol{K}]\blacklozenge\neg\phi$ is consistent. As we argued above this is the right prediction.

In addition, we also predict the following inferences.

Result 6:

- $[\boldsymbol{B}]\blacklozenge\phi \vDash [\boldsymbol{I}]\blacklozenge\phi$

By Result 4, $\mathbf{[}\boldsymbol{I}\mathbf{]}\blacklozenge\phi$ is equivalent to $\langle \boldsymbol{I} \rangle\phi$. Let $\langle \boldsymbol{I} \rangle$ correspond to the verb “consider it to be possible” in natural language. If an agent believes ϕ to be possible, then it implies that she has a piece of information of ϕ and so considers ϕ to be possible.

But the entailment from *believe-might* to *know-might* does not hold because of the presupposition.

- $[\boldsymbol{B}]\blacklozenge\phi \nvDash [\boldsymbol{K}]\blacklozenge\phi$

As argued, *know-might* sentences require the embedded *might*-sentences to be assertable. But there seems no such requirement for *believe-might* sentences. So in BSEL, belief of an epistemic possibility do not guarantee knowledge of it.

2.7.3 Negation, knowledge and belief

A crucial result in BSML is the correct prediction that free choice effects disappear in negative contexts. For example, (31-a) merely means Daiyu can't possibly be in Suzhou or Jinling.

(31) a. Daiyu might not be in Suzhou or Jinling.
b. $\rightsquigarrow$ Daiyu might not be in either cities.

Our analysis of $\blacklozenge$ can also capture this behaviour of free choice under negation.

Result 7:

- *Negation*
  $[\neg\blacklozenge(\phi \vee \psi)]^+ \vDash \neg\blacklozenge\phi \wedge \neg\blacklozenge\psi$

Proof: Suppose $\mathcal{M}, s \vDash [\neg\blacklozenge(\phi \vee \psi)]^+$, which means $\mathcal{M}, s \vDash \neg\blacklozenge(([\phi]^+ \wedge \mathrm{NE}) \vee ([\psi]^+ \wedge \mathrm{NE}))$ and $s \neq \varnothing$. This implies that $\mathcal{M}, s \dashv \blacklozenge(([\phi]^+ \wedge \mathrm{NE}) \vee ([\psi]^+ \wedge \mathrm{NE}))$, which means $s \dashv ([\phi]^+ \wedge \mathrm{NE}) \vee ([\psi]^+ \wedge \mathrm{NE})$. So $s \dashv [\phi]^+ \wedge \mathrm{NE}$ and $s \dashv [\psi]^+ \wedge \mathrm{NE}$. This means that $\exists t_1, t_2 : t_1 \cup t_2 = s$, $t_1 \dashv [\phi]^+$ and $t_2 \dashv \mathrm{NE}$. By Fact 9, $t_1 \dashv \phi$. Also there are t_3, t_4: $t_3 \cup t_4 = s$, such that $t_3 \dashv [\psi]^+$ and $t_4 \dashv \mathrm{NE}$. By Fact 9, $t_3 \dashv \psi$. So $t_2, t_4 = \varnothing$, and since $s \neq \varnothing$, $t_1 = t_3 = s$. This implies that $s \dashv \phi$ and $s \dashv \psi$, namely $s \dashv \blacklozenge\phi$ and $s \dashv \blacklozenge\psi$. Therefore, $\mathcal{M}, s \vDash \neg\blacklozenge\phi$ and $\mathcal{M}, s \vDash \neg\blacklozenge\psi$.

□

In our analysis of knowledge, we capture the factivity of knowledge using presupposition. Because presuppositions are preserved under negation, we correctly predict the factive inferences under negation.

Result 8:

- *Factivity of knowledge*
  $[\boldsymbol{K}]\phi \vDash \phi$, *in particular* $[\boldsymbol{K}]\blacklozenge\phi \vDash \blacklozenge\phi$
- *Factivity under negation*
  $\neg[\boldsymbol{K}]\phi \vDash \phi$, *in particular* $\neg[\boldsymbol{K}]\blacklozenge\phi \vDash \blacklozenge\phi$

Interestingly, if we define the anti-supported condition of $[\boldsymbol{B}]$ as follow:

$$M, s \dashv [\boldsymbol{B}]\phi \quad \text{iff} \quad \forall w \in s, \subseteq [w]_{dox} \dashv \phi$$

we will lose the duality of $[\boldsymbol{B}]$ but capture the phenomenon so-called *Neg-raising* (see Horn (1978) etc.) for believe sentences in linguistic literature, in which negation taking wide scope with respect to the *believe*-operator is interpreted as a lower scope negation:

(32) a. Daiyu does not believe that it is raining
b. $\rightsquigarrow$ Daiyu believes that it is not raining

In particular, this is preserved when the embedded clause is a *might*-sentence.

(33) a. Daiyu does not believe that it might be raining
b. $\rightsquigarrow$ Daiyu believes that it might not be raining

2.8 Discussion: *know* vs *believe*

In this section, we discuss how the agents' attitudes influence the usage of *know* and *believe* in natural language. In our proposal, *know* and *believe* differ in terms of plausibility and presupposition. And hence we can distinguish between scenarios where we should use *know* and in which we should use *believe*. Consider the following sentences in which the usage of *know* is infelicitous.

(34) a. Daiyu believes that Tsinghua is the best university in China, but she is not sure.
b. # Daiyu knows that Tsinghua is the best university in China, but she is not sure.

(35) a. Daiyu believes that Tsinghua is the best university in China and she is sure about this. But I disagree.
b. # Daiyu knows that Tsinghua is the best university in China but she is not sure. While I disagree with her.

Belief in our framework is modeled with respect to plausibility ordering. The agent always believes what is most plausible. In this sense, we argue that belief is weak because the agent does not need to have full confidence in what she believes. So in (34-a), we can use *believe* to express Daiyu's attitude towards Tsinghua. But if we say "Daiyu knows" which implies that Daiyu is sure about this matter, it will be inconsistent with her uncertain attitude. So know cannot be used in such a situation.

In (35), the speaker disagrees that Tsinghua is the best university in China, which means the presupposition of "Tsinghua is the best university in China." does not hold in the speaker's information state. Therefore, in her opinion, Daiyu's attitude towards this matter can only be a belief rather than knowledge, so she cannot express it with *know*.

This distinction also appears when embedding *might*-sentences under *know* or *believe*. If the embedded *might*-sentence cannot be asserted by the speaker, then the *know-might* sentence cannot be accepted. This is clearly illustrated by the infelicity of the following sentences where the embedded might-sentences state some objective falsity:

(36) a. # Daiyu knows that the square root of 4 might be 3 or -3.
     b. # Daiyu knows that -5 might be less than -6.

In both (36-a) and (36-b), the *might*-sentences state incorrect mathematical statements. We reject such *might*-sentences, so the two *know-might* statements give rise to a presupposition failure. In such cases, we need to use *believe* rather than *know*.

However *might*-sentences do not always express objective content. For example, in what sense can we assert the following sentences?

(37) a. ? I know that Tsinghua might be the best university in China.
     b. ? Daiyu knows that Tsinghua might be the best university in China.

This may be a complicated example. Because the assertability of the *might*-sentence "Tsinghua might be the best university in China." depends on who affirms it. In (37), there are two agents: the speaker and the subject (Daiyu). So their attitudes towards the *might*-sentence determine if (37-a) and (37-b) can be asserted. Nevertheless, the conditions of asserting the two sentences are different. In (37-a), the subject is the speaker herself, so we may need interpretation from a first person perspective. The only condition of judging this sentence is whether the speaker accepts the *might*-sentence. Whereas, in (37-b), the speaker and the subject are different. So we have to adopt a third person interpretation.

We need to check if they both agree on what the *might*-sentence expresses. To capture the contrast between first person and third person interpretation, we need to extend BSEL to a multi-agent setting (see Yan et al. (2023)).

There will be many issues arising from the interaction of agency with *know-might* and *believe-might sentences*. In this chapter, we have discussed only one small issue in two-agents (speaker-subject) situations. In Chapter 3, we provide a possible way of multi-agent extension of BSEL. This also motivates us to further explore the relationship between BSEL and classical epistemic logic.

2.9 Summary

In this chapter, we investigated epistemic *might* under knowledge and belief. We observed how WEM influences the interactions. Namely some non-classical properties of *might*, like epistemic contradictions and giving rise to free choice effects, would create tensions with classical properties of knowledge and belief, such as monotonicity, factivity of knowledge and the epistemic entailment thesis. We presented two puzzles to show these tensions, and in this way we pointed out the difficulties of studying epistemic *might* under knowledge and belief from both a technical and a philosophical perspective. The solutions to the puzzles include our consideration of how to characterize and formalize these epistemic modalities. We finally gave a formal account that captures the behaviours of their interactions.

In this chapter, we combined the following ingredients to analyze the integrating phenomena.

(i) We adopted an expressivist analysis for epistemic *might*. We assumed the non-classical properties of *might*: epistemic contradictions and non-factivity. Also we considered the observations from linguistic literature that epistemic *might* would give rise to free choice and ignorance effects.

(ii) We employed BSML as the baseline framework to deal with the pragmatic inferences. We further extended BSML to an epistemic framework BSEL to analyze the epistemic modals within a team-based framework.

(iii) To extend BSML, we used plausibility models of epistemic logic to formalize belief. The motivation is to distinguish belief and knowledge in the sense that the interpretation of belief should quantify over a smaller set of worlds than knowledge. This guarantees the entailment from knowledge to belief for factual propositions but not for *might* sentences.

(iv) Finally we argued that knowledge should have the following properties: a. factivity, even for the *might*-sentences; b. projection of factivity under negation. We found that the classical **S**5 modality for knowledge failed to capture these features. So we proposed a novel semantics for knowledge in terms of presupposition.

We argued that epistemic possibilities can be knowledgeable, provided *might*-sentences can be assertable. Determining whether a might sentence is assertable depends on the information state of the speaker and the subject as well as their interactions. We discussed this issue in terms of the contrast between *know* and *believe* in BSEL. In the next chapter, we will consider more details of multi-agency.

CHAPTER 3 SHIFTING PERSPECTIVES: A MULTI-AGENT EXTENSION OF BSEL

Epistemic *might* is argued to be perspective sensitive (see Speas and Tenny (2003); Stephenson (2007); Egan et al. (2005); MacFarlane (2011) among others). The truth or falsity of a *might* claim depends on the perspective under which it is made. Consider the sentence in (1):

(1) It might be raining.

Based on the classical possible worlds analysis for *might*, an epistemic reading of (1) is that in **some** worlds that stand in an epistemic relation to the actual world (Hintikka (1962); Kripke (1963)) or compatible with what is known in the actual world (Kratzer (1977); Veltman (1996)), it is raining. As MacFarlane noted, these theories all propose a semantics that presupposes a knower(s) whose knowledge (or epistemic state) is included as a parameter in the circumstances of evaluation. And epistemic *might* is treated as an **existential** quantifier over the possibilities within her (their) knowledge. Perpectives are represented by the epistemic states of the knower(s). Then a question arises: whose knowledge is being expressed? In this chapter, we tackle this issue by extending the scope of BSEL to accommodate multi-agent environments, and explore the interactions between those epistemic modalities in multi-agent situations.

3.1 Introduction

In Chapter 2, we assumed that a plain epistemic *might*-claim, such like (1), expresses the information that the speaker has. We proposed a framework, known as Bilateral State-based Epistemic Logic (BSEL), which provides a non-classical analysis for epistemic *might*, wherein a *might*-claim is interpreted as the proposition it takes being consistent with the speaker's information. BSEL is like a variant of assertability logic, which interprets formulas as either assertable or rejectable (rather than true or false) relative to an information state that is defined as a subset of possible worlds. In BSEL models, an information state s, which serves as the current point of evaluation, is always assumed to model the information held by the speaker and thus reflects their perspective. The assertability of a *might*-formula $\blacklozenge\phi$ is defined such that there exists a non-empty substate of s in which ϕ is assertable. This definition effectively implies that the *might*-claim is assertable within the speaker's information. In BSEL we argue that a plain *might*-claim is always linked to the speaker's perspective.

In addition to the properties of epistemic *might* discussed in Chapter 2, another notable feature of this modal pertains to its perspective-sensitivity. This property is established as evidenced by its parallel behaviours with predicates of personal taste (e.g., Stephenson (2007)). For example, it is argued that the *might*-claim embedded in belief reports is linked to the immediate subject in the sentence, such that the *might*-claim in an utterance like "Daiyu believes it might be raining." expresses an attitude from Daiyu's perspective, specifically reporting Daiyu's belief. BSEL can capture this phenomenon by virtue of the semantics for epistemic *might* providing a locality effect. That is, *might* operates directly on the local information state that represents Daiyu's doxastic state. BSEL generates Daiyu's doxastic state from the information state s through accessibility and plausibility relations, allowing the speaker's perspective to play a role in interpreting the subject's attitude. This formalization thus enables the representation of both the speaker's and the subject's perspectives.

Furthermore, the formalization of the *might*-claim embedded in knowledge is achievable with BSEL. The factivity of knowledge implies that the assertability of an utterance such as "Daiyu knows it might be raining." requires the *might*-claim to be assertable with respect to both the current information state s and Daiyu's information. To address this issue, we proposed a new definition of knowledge in BSEL that accounts for this requirement by forcing that the embedded *might* also operates on s, in addition to operating on Daiyu's information if the speaker is not Daiyu herself. As previously indicated, we assume that s represents the speaker's perspective, which leads us to argue that both the speaker and the subject serve as the knowers of the *might*-claim embedded under the verb *know*.

However, it should be noted that BSEL models are not inherently designed for multi-agent contexts, in that there is only one accessibility and plausibility relation present within the model. As such, the model can formalize only one agent's beliefs and knowledge. We can employ BSEL to formalize the perspective-sensitivity of epistemic *might* in certain cases where there is only one knower involved. Specifically, this includes situations where the interpretation of the proposition underlying the *might*-claim only requires one agent's perspective, such as in the case of plain *might*-claims or believe-*might* sentences that contain only one subject. And we can also formalize know-*might* sentences that contain only one subject, where both the speaker's and the subject's perspectives are involved in the interpretation by virtue of the definition of knowledge. But it struggles to capture cases that require more than one agent's belief and knowledge within the framework of BSEL. Consider the following example:

(2) a. Daiyu believes that Baoyu believes that it is raining.
b. I know it is not raining but Baoyu believes that it is raining.

As a consequence, the scope of our analysis with BSEL models is not equipped to model situations where multiple knowers of a *might*-claim with different perspectives may be present. For example, the cases of disagreement pose a significant challenge in identifying a *might*-claim that is reasonable for both the speaker to assert and for a listener to dispute (see MacFarlane (2011); Von Fintel and Gillies (2008)). However reasoning about epistemic "might" may differ in multi-agent scenarios compared to classical single situations. Consider, the following example:

(3) a. Baoyu says: "It might be raining."
b. Daiyu says: "It is not raining."

In the given scenario, it appears that there is no epistemic contradiction. This is because the conjunction of the two statements should be construed as "Baoyu believes that it might be raining and Daiyu believes that it is not raining." rather than "it might be raining and it is not raining." Consequently, the modal *might* in the first statement is interpreted solely in relation to Baoyu's epistemic state.

To express and reason about epistemic *might* in multi-agent scenarios, this chapter aims to expand the scope of BSEL beyond the single-informant case and provide a framework for investigating epistemic *might* and its interactions with knowledge and belief in multi-agent environments. By analyzing the perspective-sensitivity of *might*, we can better understand the nature of knowledge and belief in social interactions.

We propose an extension to BSEL by incorporating a two-dimensional semantics of Epistemic Friendship Logic (EFL, Seligman et al. (2011, 2013)). EFL possesses mechanisms in the models that make it possible for transformations between different agents and thus can capture perspective shifts. We leverage this capability to formalize the concept of perspective sensitivity.

This chapter is structured in the following manner: In Section 3.2, we present arguments for the perspective-sensitivity of epistemic *might* by examining its parallel behavior with predicates of personal taste. In Section 3.3, we introduce a two-dimensional semantics of Epistemic Friendship Logic. In Section 3.4, we make our initial attempt to combine the EFL and BSEL models to capture the perspective-sensitivity of epistemic *might* in multi-agent contexts. In Section 3.5, we further extend the multi-agent model by incorporating belief and knowledge. Finally, in Section 3.6, we provide a comprehensive overview of the model and present some results obtained from the fusion of the two logics.

3.2 Perspective-sensitivity of epistemic *might*

One compelling line of evidence that supports the notion of epistemic *might* being sensitive to the perspectives of agents is its parallel behavior with predicates of personal taste, see Lasersohn (2005); Stephenson (2007) among others. In this section, we start by examining these parallel behaviors. Then, we will explore what factors need to be taken into account in the semantics to capture the perspective sensitivity of epistemic *might*.

3.2.1 Epistemic *might* and predicates of personal taste

Predicates of personal taste refer to expressions or judgments that reflect a person's individual preferences or opinions, such as "This girl is **beautiful**" or "The cake is **tasty**". So they are typically considered to be perspective-sensitive. This implies that the interpretation of such statements is contingent upon the perspective or standpoint of the individual making the statements. For example, the statement "This ice cream is delicious." involves a predicate of personal taste—*delicious*. Whether or not the ice cream is deemed delicious will hinge on the individual's unique tastes and preferences.

One parallel between epistemic *might* and predicates of personal taste is that both involve a degree of subjectivity. In the case of epistemic *might*, the strength of one's knowledge or belief may depend on factors such as the quality and reliability of the information available, and one's own cognitive abilities and biases. Similarly, predicates of personal taste are influenced by the individual's internal state. The parallel behaviours can be observed in the cases of disagreement and attitude contexts.

3.2.1.1 Disagreement

Consider the following two different scenarios, and answer the questions.

> **Scenario 1**: Daiyu does not like food with a special smell. At a party, she hears Liulaolao say that "Garlic is delicious!". Daiyu doesn't like garlic. Question: Can Daiyu disagree with Liulaolao by asserting "No, garlic is not delicious"?
>
> **Scenario 2**: Scene as before. Liulaolao says "Garlic is delicious to me!" Daiyu doesn't like garlic. Question: Can Daiyu disagree with Liulaolao by asserting "No, garlic is not delicious"?

In Scenario 1, it is acceptable for Daiyu to reject Liulaolao, but not in Scenario 2. Thus, we can argue that disagreement is possible in Scenario 1, meaning that the phrase "No, garlic is not delicious" can be asserted in the dialogue. However, in Scenario 2, the disagreement is problematic as it is inappropriate to use this phrase to oppose the assertion "Garlic is delicious to Liulaolao" which express Liulaolao's own preference.

Our intuition in Scenario 1 is that Liulaolao and Daiyu judge the proposition "Garlic is delicious ($\mathtt{Delic}$(g))" from their own perspective. Liulaolao judges the proposition to be true or assertable to her, while Daiyu judges it to be false or unassertable. The disagreement between them expresses their respective judgments.

Agent	Judge: $\mathtt{Delic}$(g)	Outcome
Liulaolao	Yes	Garlic is delicious
Daiyu	No	Garlic is not delicious

In Scenario 2, it is no longer a simple matter of personal taste and the judgment of the predicate "delicious" as being solely based on personal preference. Instead, the interpretation of the word "delicious" is based on Liulaolao's perspective ($\mathtt{Delic}_{Liu}$(g)). Thus, Daiyu's opposition from her own perspective seems infelicitous[1].

Agent	Judge: $\mathtt{Delic}_{Liu}$(g)	Outcome
Liulaolao	Yes	Garlic is delicious to me
Daiyu	-	-

The behaviour of epistemic *might* is similar. Consider the examples and questions in the following scenarios[2].

1 In fact, Daiyu could hardly judge this sentence accurately because she could not fully understand Liulaolao's perspective to determine if garlic is delicious.

2 The original version of this example comes from Kratzer (1986); MacFarlane (2011).

Scenario 3: You are in the Bamboo Lodge. You overhear Xiren and Qingwen talking outside the door. Xiren says “Baoyu might be in the Bamboo Lodge”. You have never left the Bamboo Lodge, neither seen Baoyu there. You know that Baoyu is not in the Bamboo Lodge. Question: Do you think that Xiren speaks falsely?

Scenario 4: Scene as before. Xiren says “I don’t know anything that would rule out Baoyu’s being in the Bamboo Lodge ” (or “For what I know, Baoyu is in the Bamboo Lodge”). You have never left the Bamboo Lodge, neither seen Baoyu there. You know that Baoyu is not in the Bamboo Lodge. Question: Do you think that Xiren speaks falsely?

Similarly, we can answer “Yes” to the question in Scenario 3 and “No” to the question in Scenario 4. This is because in the case of Scenario 4, Xiren is only sharing information based on her own knowledge and expressing a *might*-claim from her own perspective. However, in the case of Scenario 3, it is acceptable to reject Xiren’s assertion as both parties are providing their evaluation of the proposition “Baoyu might be in the Bamboo Lodge ($\texttt{Might}$(b))” based on their available information. Suppose you are Daiyu, then we can conclude by the following tables.

Agent	Judge:$\texttt{Might}$(b)	Outcome
Xiren	Yes	Baoyu might be in the Bamboo Lodge
Daiyu	No	Baoyu is not in the Bamboo Lodge

Agent	Judge: $\texttt{Might}_{Xi}$(b)	Outcome
Xiren	True	For what I know, Baoyu might be in the Bamboo Lodge
Daiyu	-	-

3.2.1.2 Attitude reports

An epistemic agent (knower, as mentioned above) is presupposed when expressing claims that contain epistemic *might*. In attitude reports, it is clear to figure out the relevant agent(s) of *might*-claims and predicates of personal. Consider the following examples in which the expressions are embedded under the attitude verb *believe*.

(4) a. The eggplant is tasty.

b. Liulaolao believes that the eggplant is tasty.

(5) a. It might be an eggplant.
 b. Liulaolao believes that it might be an eggplant.

In (4-b), the predicate “tasty” of personal taste is related to the subject “Liulaolao”. This statement expresses Liulaolao’s internal state of rating the eggplant dish. Similarly, the statement in (5-b) reports Liulaolao’s belief of an epistemic possibility, as indicated by the epistemic reading of (5-a).

Likewise, the statement in (5-b) reports Liulaolao’s belief of an epistemic possibility, provided the epistemic reading of (5-a). The embedded *might*-claim expresses the possibilities of Liulaolao’s epistemic state. (5-b) is true if and only if Liulaolao’s belief excludes the possibility that it is not an eggplant. Therefore holder of the belief attitude is the epistemic agent (knower) of the *might*-claim.

Therefore, when a predicate of personal taste or a *might*-claim is embedded under another verb, the predicate and the *might* describe the mental state of the immediate subject.

Now consider the following more complex sentences, which show a hierarchical relationship between the clauses.

(6) a. Daiyu believes that Liulaolao believes that the eggplant is tasty.
 b. Daiyu believes that Liulaolao believes that it might be an eggplant.

The proposition expressed in (6-a) communicates that the eggplant being delicious is solely a preference held by Liulaolao. Indeed, even though Daiyu acknowledges that Liulaolao has this particular preference, her personal stance on the deliciousness of the eggplant is not conveyed by the sentence. Thus, in cases of multiple embeddings, such as in (6-a), the predicate of personal taste is semantically linked to the immediate subject of the subordinate clause. In (6-a), it relates exclusively to Liulaolao’s perspective.

In a parallel manner, (6-b) says that Daiyu’s belief is presenting a report about Liulaolao’s belief where it does not exclude the possibilities of the dish being an eggplant. However, it refrains from indicating whether Daiyu herself subscribes to this possibility. Thus, while Daiyu reports on Liulaolao’s belief, her own stand on the proposition remains undisclosed in the context of (6-b).

In summary, similar to the predicates of personal taste, a *might*-claim in attitude reports is always linked to the perspective of its immediate subject. This holds regardless of whether multiple embeddings occur. The immediate subject is the knower(s) of the embedded *might*-claims. An exception arises when a *might*-claim is embedded under the factual verb *know*, a case which will be further discussed in Section 3.4.

3.2.2 The role of agency in the semantics of *might*

The parallels highlight the perspective-sensitivity of epistemic *might*. The perspectives of the relevant agents affect the interpretation and understanding of *might*-claims. Additionally, it suggests that the semantics of epistemic *might* should include a "taste parameter" (Lasersohn (2005)) which formalizes the perspective of an agent or a group of agents (see Egan et al. (2005); Stephenson (2007); MacFarlane (2014) among others). This theory is also known as relativism in philosophy and semantics.

Notice that the study of relativist analysis for epistemic modals presents its own set of difficulties and obstacles (see Stojanovic (2007)). This chapter does not delve into the philosophical discussions surrounding relativism[1]. In Chapter 2, BSEL was proposed as a means to offer an expressivist analysis for the epistemic *might*. Expressivism also emphasizes the role of agency in a way that epistemic modals are considered to express a property of the agent's attitudes. This theory departs from the approach of relativism, however, the idea of adding the taste parameter inspires us to incorporate agent factors into BSEL's analysis of *might* to capture the perspective sensitivity explicitly.

As noted in Section 1, BSEL models provide a limited means of capturing the perspective-sensitivity of the epistemic *might*. In BSEL, the speaker is assumed to be the relevant agent the information state specifies. So formulas are always interpreted relative to the speaker's perspective. BSEL models encounter two problems in multi-agent scenarios.

- The evaluation parameter only contains one information state, representing the perspective of one speaker, and one accessibility relation, representing the information state of one subject.
- Factors of agency are not represented explicitly in the semantics.

The first issue arises from the fact that BSEL models cannot express multiple agents. Thus it falls short in capturing sentences that involve more than one agents and the inferences of these sentences. For example, BSEl cannot formalize the following reasoning:

1 There is already a gamut of debates between contextualist and relativist treatment of epistemic modals revolves around the question that how to determine the epistemic agents in evaluations, see DeRose (1991); Egan et al. (2005); Stojanovic (2007); MacFarlane (2011) among others. Such dispute is not our concern in this chapter. We only focus on the formalization of epistemic agents in the state-based account for *might* proposed in BSEL.

(7) a. Daiyu believes that Baoyu knows that Liulaolao knows that it is raining.
b. $\Rightarrow$ It is raining.

At most, BSEL models can capture the scenario involving only two agents, namely one speaker and one subject. For example,

(8) Daiyu knows that it might be raining.

This statement involves two agents, the speaker and Daiyu, if Daiyu herself is not the speaker. The speaker's perspective is represented by the current information state of evaluation. Daiyu's perspective is represented by the possible worlds generated by accessibility relation. A BSEL model that demonstrates the assertability of the statement is presented in the Figure 3.1.

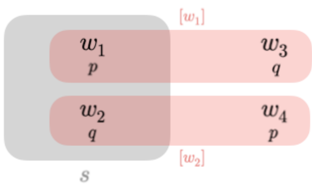


Figure 3.1 A simple case of two agents

Let p be the proposition "it is raining" and $[w]$ be the equivalence class of w. The grey region in the model depicts the current information state that represents the speaker's perspective, while the pink region represents Daiyu's epistemic states generated from possible worlds in s via the equivalence relation, which consists of $[w_1]$ and $[w_2]$. There exists a p-world (w_1) in s, so s supports $\blacklozenge p$. Moreover, because each $[w_1]$ and $[w_2]$ contains a p-world (w_1/w_4), proposition p is assertable from Daiyu's perspective. Based on the definition of knowledge in BSEL, the statement in (8) is assertable by the model.

However, as the model of BSEL includes only a single accessibility relation, this approach is unable to differentiate between the epistemic states of various subjects. Should an additional agent be introduced within the sentence, this model would fail to differentiate it from Daiyu, thereby multiple agents cannot be represented.

The second issue concerns the implicit expression of agency through information states in BSEL. As argued, the agency plays an important role in understanding and interpreting *might*-claims. However, in BSEL the current information state, as the evaluation parameter, is assumed to represent the speaker's perspective. But we do not explicitly define this in the model. As a result, our representation of perspective-sensitivity through the model could potentially be viewed as an overinterpretation, rather than an intrinsic characteristic of the model. Therefore, the role of agency necessitates an explicit definition to provide a further application.

Our objective is to formalize multi-agent scenarios while preserving the properties and interactions of those epistemic modalities as discussed in Chapter 2. In order to accomplish this, we must address and overcome the two issues intrinsic to BSEL models. The proposed solution is to incorporate factors of agency into the BSEL model.

3.3 Epistemic friendship logic

In this section, we introduce a two-dimensional semantics for Epistemic Friendship Logic (EFL) proposed in Seligman et al. (2011), and show how it works for shifting perspectives.

3.3.1 Two-dimensional semantics for EFL

In Seligman et al. (2011, 2013), a two-dimensional modal logic was proposed for analyzing and reasoning knowledge and social relationships in social networks. In this logic, agents are placed in the model, and formulas are evaluated with respect to both possible worlds and agents. So it allows us to express aspects of agency. In this chapter, we only focus on the static fragment of EFL.

The language $\mathcal{L}_{EFL}$ is based on atoms of two types: propositional variables $p \in Prop$ and (a finite set of) agent nominals $n \in ANom$. The propositional variables can represent the indexical propositions like "I am hungry," and the nominals name the agents.

Definition 17 (Language $\mathcal{L}_{EFL}$(Seligman et al. (2013))**):**

$$\varphi ::= p \mid n \mid \neg\varphi \mid \varphi \wedge \varphi \mid K\varphi \mid F\varphi \mid A\varphi$$

The operator K stands for knowledge. We read the formula $F\phi$ as "all my friends have the property ϕ" and the formula $\mathbf{A}\phi$ as "every agent has the property ϕ."

Models for $\mathcal{L}_{EFL}$ are Kripke-style models of the form $\mathcal{M}_{EFL}$=$\langle W, A, R, f, V\rangle$.

- W is a set of possible worlds, and A is a set of agents.
- R is a family of accessibility relations, which is reflexive, symmetric and transitive. R_a is the epistemic indistinguishability relation for each agent $a \in A$.
- f is a family of symmetric and irreflexive relations. f_w for each $w \in W$ represents the friendship relation in state w.
- V is a valuation function: mapping propositional variables $p \in Prop$ to subsets of

$W \times A$, namely $(w, a) \in V(p)$. It means that the proposition p holds for agent a in world w.

It defines g to be an assignment function on $\mathcal{M}_{EFL}$ mapping each agent nominal $n \in ANom$ to the agent $g(n) \in A$ named by n.

This model is a two-dimensional model. A formula, in classical modal logic, is satisfied by exactly one evaluation index, while in the case of EFL models, would have to be the pair (w, a).

Definition 18 (Semantics (Seligman et al. (2013))): *The EFL models are used to interpret $\mathcal{L}_{EFL}$ in a double-indexical way:*

$$
\begin{array}{lll}
\mathcal{M}_{EFL}, w, a \vDash p & \textit{iff} & (w, a) \in V(p), \textit{ for } p \in Prop \\
\mathcal{M}_{EFL}, w, a \vDash n & \textit{iff} & g(n) = a \textit{ for } n \in ANom \\
\mathcal{M}_{EFL}, w, a \vDash \neg\phi & \textit{iff} & \mathcal{M}_{EFL}, w, a \nvDash \phi \\
\mathcal{M}_{EFL}, w, a \vDash \phi \wedge \psi & \textit{iff} & \mathcal{M}_{EFL}, w, a \vDash \phi \textit{ and } \mathcal{M}_{EFL}, w, a \vDash \psi \\
\mathcal{M}_{EFL}, w, a \vDash K\phi & \textit{iff} & \mathcal{M}_{EFL}, v, a \vDash \phi \textit{ for every } v \in W \textit{ s.t. } R_a(w, v) \\
\mathcal{M}_{EFL}, w, a \vDash F\phi & \textit{iff} & \mathcal{M}_{EFL}, w, b \vDash \phi \textit{ for every } b \in A \textit{ s.t. } f_w(a, b) \\
\mathcal{M}_{EFL}, w, a \vDash A\phi & \textit{iff} & \mathcal{M}_{EFL}, w, b \vDash \phi \textit{ for every } b \in A
\end{array}
$$

A proportional letter p is true with respect to indexes w, a if and only if p holds at w for agent a. A nominal n is true with respect to w, a if and only if n is the name of the agent a. $K\phi$ is true at w for a if and only if for agent a, ϕ holds at every accessible worlds v from w in term of relation R_a. The truth of $F\phi$ at w for agent a means that ϕ holds at w for every agent b who has friendship relation with a, namely every friend of a. The formula $\mathbf{A}\phi$ is true at w for a if and only if ϕ holds at w for every agent in A.

To illustrate how this semantics works, consider the following toy example.

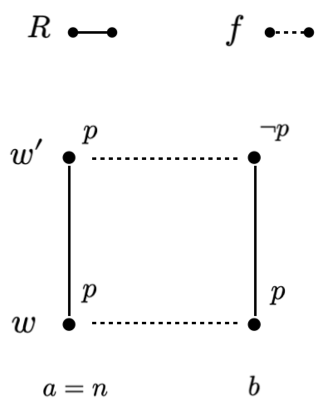


Figure 3.2 EFL model

In the model depicted by Figure 3.2, $W = \{w, w'\}$ and $A = \{a, b\}$. The agent a has a name n, viz $g(n) = a$. The solid line represents the equivalence relation R and the dotted line stands for the friendship relation f. As the model shows, $R_{a,b} = \{w, w'\}$, and at each world agents a and b are friends. $V(p) = \{(w, a), (w', a), (w, b)\}$. In this model, we can get the following results:

- $\mathcal{M}_{EFL}, w, a \models p$
- $\mathcal{M}_{EFL}, w, a \models n$
- $\mathcal{M}_{EFL}, w, a \models Kp$
- $\mathcal{M}_{EFL}, w, b \models \neg Kp$
- $\mathcal{M}_{EFL}, w, a \models Fp$
- $\mathcal{M}_{EFL}, w', a \models \neg Fp$
- $\mathcal{M}_{EFL}, w, a \models \neg FKp$
- $\mathcal{M}_{EFL}, w, a \models \neg KFp$

It is different between the ordering of K and F. Let p be the proposition 'I am hungry'. Then KFp says " I know that all my friends are hungry" and FKp means " Each of my friends knows that she is hungry". The existential duals of modal operators are defined as usual: $\langle K \rangle = \neg K \neg$; $\langle F \rangle = \neg F \neg$; $\langle \mathbf{A} \rangle = \neg \mathbf{A} \neg$.

Agents nominals play an important role in EFL. Assuming the rigid reading, every nominal n specifies an agent. This is inspired by hybrid logic (see Blackburn and Seligman (1995); Seligman (1997) etc, and Areces and ten Cate (2007) provides a good overview) which uses nominals name possible worlds. In addition, another hybrid-style operator @ can be defined as a derived one: $@_n\phi = \langle \mathbf{A} \rangle (n \wedge \phi)$, which means "For n, ϕ holds".

$$\mathcal{M}_{EFL}, w, a \models @_n\phi \quad \text{iff} \quad \mathcal{M}_{EFL}, w, g(n) \models \phi$$

The formula $@_n\phi$ is true relative to indexes w and a if and only if ϕ is true relative to the reference of n. Notice that the point a in fact does not matter in the truth-condition for $@_n\phi$. Since the operator $@_n$ moves the current index of agent to the reference of n whatever a is.

Therefore, the operator $@_n$ allows for a shift in perspective, in the way that the interpretation of $@_n\phi$ requires a move to the agent named n from the current agent. For instance, let p represent the proposition "The cake is delicious" and n represent Daiyu. The formula $@_np$ is true iff $(w, g(n)) \in V(p)$, which means "The cake is delicious" is true for Daiyu at world w. In this evaluation, Daiyu is the speaker of p and "delicious" is linked to her. The perspective can be changed by switching the operator $@_n$. If the nominal m is Baoyu, then $@_mp$ means the cake is delicious for Baoyu and the predicate "delicious" would be linked to Baoyu.

$\mathcal{M}_{EFL}$ is a named agent model, if every agent in $\mathcal{M}_{EFL}$ has a name, viz, for each

$a \in A$, there is a nominal $n \in ANom$ such that $g(n) = a$. By assuming a named agent model[1], another hybrid-style operator can be defined in EFL, $\downarrow n$, which provides a way of indexically referring to the current agent.

$$\downarrow n\, \phi = \bigvee_{m \in ANom} (m \wedge \phi^n_m)$$

The semantics for $\downarrow n$ is defined as follows.

$$\mathcal{M}_{EFL}, w, a \models \downarrow n\, \phi \quad \text{iff} \quad \mathcal{M}^n_a, w, a \models \phi$$

where $\mathcal{M}^n_a$ is the result of changing $\mathcal{M}_{EFL}$ by forcing n to name a. More precisely, the assingment function g defined on $\mathcal{M}_{EFL}$ now is changed to g^n_a, and g^n_a(m)=a if $m = n$ and $g(m)$ otherwise. The operator $\downarrow n$ behaves like the $\downarrow$ binder in hybrid logic which binds a nominal to the point of evaluation. In EFL, $\downarrow n\, \phi$ is true relative to w and a if and only if ϕ is true relative to w and a when n refers to a.

This semantics allows us to express indexicals under knowledge in EFL. For example, let p be the propersition "I'm hungry" and nominal m be Daiyu.

(9) $\downarrow n @_m K @_n p$: Daiyu knows that I am hungry.

(10) $\downarrow n F K \langle F \rangle n$: All my friends know they are friends with me.

The operator $\downarrow n$ binds the nominal n that occurs in the bound formula, which force n refer back to the current agent in the pair of evaluation. In this way, "I" and "me" represented by n refers to the speaker rather than the subject of the sentence. If we don't formalize the sentence in (9) with $\downarrow n$, for example, then it will be translated as $@_m K p$ in which "I" refers to "Daiyu" instead of the speaker.

In this section, we have reviewed a two-dimensional semantics for Epistemic Friendship Logic, in which agents are explicitly placed in the model. Formulas are interpreted relative to a pair of possible world w and agent a, rather than a single point of evaluation. This provides a different approach to multi-agent modeling than classical multi-agent epistemic logic. The nominals and operators $@_n$ and $\downarrow n$, borrowed from hybrid logic, enhance the expressivity of the logic. Another important benefit is that it can capture the friendship relationship between agents and express knowledge in social networks. However, this is not the focus of this chapter. Our goal is to extend logic BSEL by taking EFL's approach to modeling multi-agent scenarios. For further discussion on EFL and its applications, see references such as Seligman et al. (2011); Ruan and Thielscher (2011); Liu et al. (2014).

1 Otherwise, the operator can be introduced as a primary one.

In the next section, we attempt to integrate this two-dimensional framework with the BSEL fragment that lacks knowledge and belief, in order to address the perspective-sensitivity of epistemic *might*.

3.4 Multi-agent extension of BSEL: first attempt

In this section, we extend BSEL using the semantics of EFL introduced in the previous section. Initially, we only extend the fragment of BSEL containing the epistemic *might* represented by the operator $\blacklozenge$. In following sections, we will consider knowledge and belief.

We start with a simple language $\mathcal{L}$ which is a combination of fragments of EFL and BSEL.

Definition 19 (Language $\mathcal{L}$)**:** *The language is based on three types of atoms: propositional variables $p \in Prop$, nominols $n \in ANom$, and the non-emptiness atom* NE.

$$\varphi ::= p \mid n \mid \mathrm{NE} \mid \neg\varphi \mid \varphi \wedge \varphi \mid \varphi \vee \varphi \mid \blacklozenge\varphi \mid @_n\varphi \mid \ \downarrow n\ \varphi$$

As shown above, we only keep the non-emptiness atom and operator $\blacklozenge$ representing epistemic *might* of BSEL and the hybrid-style operators of EFL in language $\mathcal{L}$. As in EFL, we read $@_n\phi$ as "for agent named n, ϕ holds"; and $\downarrow n$ behaves like a binder.

A model of $\mathcal{L}$ is a tuple $\mathcal{M} = \langle W, A, R, g, V\rangle$. The component of $\mathcal{M}$ is like the model of EFL without the friendship relations. W is the set of possible worlds and A is the set of agents. R is a family of equivalence relations, and R_a represents the relation for agent $a \in A$. g is an assignment function: $ANom \to A$. And V is an evaluation function from Prop to the subsets of $W \times A$, with $(w, a) \in V(p)$.

It is a state-based system, viz, formulas of $\mathcal{L}$ are interpreted with respect to an information state s. To make this, it is necessary to lift the single evaluation pair (w, a) to a set. One approach to lift is: an information state s is defined as a set of indices i. An index i is a pair $i = \langle w_i, a_i\rangle$. So an information state, in fact, is a set of pairs of world w and agent a. For the sake of simplicity, we assume the agents in an information state s are the same one. Hence we have the constraint that for all $i, j \in s$, $a_i = a_j$. Because of the identity of agents in s, we label the agent as a subscript of s: s_a, highlighting the epistemic agent (knower) of the information to be a. Now we can define the semantic clauses.

Definition 20 (Semantics)**:** *Formulas of $\mathcal{L}$ are interpreted in models $\mathcal{M}$ with respect to a state $s \subseteq W \times A$. Both support $\vDash$, and anti-support $\Dashv$ conditions are specified.*

$$
\begin{array}{lll}
\mathcal{M}, s_a \vDash p & \textit{iff} & \forall i \in s_a : (w_i, a_i) \in V(p) \\
\mathcal{M}, s_a \dashv p & \textit{iff} & \forall i \in s_a : (w_i, a_i) \notin V(p) \\
\mathcal{M}, s_a \vDash n & \textit{iff} & \forall i \in s_a,\ g(n) = a \\
\mathcal{M}, s_a \dashv n & \textit{iff} & \forall i \in s_a,\ g(n) \neq a \\
\mathcal{M}, s_a \vDash \mathtt{NE} & \textit{iff} & s_a \neq \varnothing \\
\mathcal{M}, s_a \dashv \mathtt{NE} & \textit{iff} & s_a = \varnothing \\
\mathcal{M}, s_a \vDash \neg\phi & \textit{iff} & \mathcal{M}, s_a \dashv \phi \\
\mathcal{M}, s_a \dashv \neg\phi & \textit{iff} & \mathcal{M}, s_a \vDash \phi \\
\mathcal{M}, s_a \vDash \phi \vee \psi & \textit{iff} & \exists t, t' : t \cup t' = s_a \textit{ and } \mathcal{M}, t \vDash \phi \textit{ and } \mathcal{M}, t' \vDash \psi \\
\mathcal{M}, s_a \dashv \phi \vee \psi & \textit{iff} & \mathcal{M}, s_a \dashv \phi \textit{ and } \mathcal{M}, s_a \dashv \psi \\
\mathcal{M}, s_a \vDash \phi \wedge \psi & \textit{iff} & \mathcal{M}, s_a \vDash \phi \textit{ and } \mathcal{M}, s_a \vDash \psi \\
\mathcal{M}, s_a \dashv \phi \wedge \psi & \textit{iff} & \exists t, t' : t \cup t' = s_a \textit{ and } \mathcal{M}, t \dashv \phi \textit{ and } \mathcal{M}, t' \dashv \psi \\
\mathcal{M}, s_a \vDash @_n\phi & \textit{iff} & \mathcal{M}, s_a[g(n)] \vDash \phi \\
\mathcal{M}, s_a \dashv @_n\phi & \textit{iff} & \mathcal{M}, s_a[g(n)] \dashv \phi \\
\mathcal{M}, s_a \vDash \downarrow n\ \phi & \textit{iff} & \mathcal{M}^n_a, s_a \vDash \phi \\
\mathcal{M}, s_a \dashv \downarrow n\ \phi & \textit{iff} & \mathcal{M}^n_a, s_a \dashv \phi \\
\mathcal{M}, s_a \vDash \blacklozenge\phi & \textit{iff} & \textit{there is a non-empty subset } s'_a \textit{ of } s_a\textit{, such that } s'_a \vDash \phi \\
\mathcal{M}, s_a \dashv \blacklozenge\phi & \textit{iff} & \mathcal{M}, s_a \dashv \phi
\end{array}
$$

- $i[g(n)] = \langle w_i, g[n] \rangle$
- $s_a[g(n)] = \{i[g(n)] \mid i \in s\}$
- $\mathcal{M}^n_a = \langle W, A, R, g^n_a, V \rangle$, where g^n_a is a n-variant of g.

As shown in Definition 20, it is a two-dimensional state-based semantics. A propositional variable p is supported in information state s_a if and only if for every index i in s_a, p holds for agent a_i at world w_i; it is rejected by s_a if and only if p doesn't hold for each pair (w_i, a_i) in s. A nominal n is supported by s_a if and only if n is the name of the agent a who is the knower of s_a; and n is anti-supported if and only if n is the name of someone other than a. The atom NE is supported provided s_a is not empty and it is anti-supported if and only if s_a is empty.

For the semantic clauses of the connectives, it is defined in the same way of BSEL. The formula $@_n\phi$ is supported in s_a if and only if ϕ is supported in $s_a[g(n)]$ which is the information state after switching the agent a of s_a to the agent named n[1]; and the

1 It could assume that agent a is maximally informed while agent $g(n)$ is maxmally ignorant. Normally this can be captured by assigning a singleton to a's information state and the set of all possible worlds to $g(n)$'s information

formula is anti-supported provided $s_a[g(n)]$ anti-supported ϕ. The conditions of $\downarrow n$ can be reduced to the conditions of ϕ interpreted in the model $\mathcal{M}_a^n$ which is the resulting model by substituting g with g_a^n. The formula $\blacklozenge\phi$ is supported in s_a if and only if there is at least one pair of (w, a) supporting ϕ; and the formula is anti-supported if and only if ϕ is anti-supported in s_a (there is no pair of (w, a) supporting ϕ).

This semantics allows us to capture the perspective-sensitivity of epistemic *might*. Perspectives are represented by the information state that incorporates agency. The agent to whom the *might*-claim is linked can be specified by the $@_n$ operator. Shifting perspectives is to switch an agent to another one in a same information state s. For example, consider the case of disagreement of Scenario 3 and 4 in Section 3.2. Let p be the proposition “Baoyu is in the Bamboo Lodge” and $\blacklozenge p$ be the *might*-claim “Baoyu might be in the Bamboo Lodge”, which has different values from different perspectives. The claim is assertable for Xiren (x) and Qingwen (n), but rejectable for Daiyu (d). The following table shows the formalization of this scenario.

Agent	“Baoyu might be in the Bamboo Lodge”	Judgement	Formalization
Xiren (x)	$@_x\blacklozenge p$	assertable	$\mathcal{M}, s_a[g(x)] \vDash \blacklozenge p$
Qingwen (n)	$@_n\blacklozenge p$	assertable	$\mathcal{M}, s_a[g(n)] \vDash \blacklozenge p$
Daiyu (d)	$@_d\blacklozenge p$	rejectable	$\mathcal{M}, s_a[g(d)] \Dashv \blacklozenge p$

As shown in the table, a *might*-formula $\blacklozenge p$ is interpreted differently from various perspectives.

It is important to highlight that classically logical methodologies often encounter difficulties in addressing instances of **faultless disagreement**, which means a situation where two groups assert opposing views, both of which can be justified, resulting in no party being at fault. This is because it is impossible that p and $\neg p$ are both true. In conversations, it is reflected as a credible proposition that not only can the speaker legitimately assert, but also a rational listener can justifiably dispute.

state. However, we have to handle the situation differently in our model since all states have the same cardinality. The underlying idea behind this approach is that different agents have access to the same information base, but they can consider different possibilities and make different assertions. The model assumes that the information state refers to the current information possessed by the agents, and it is important to note that this does not encompass the entirety of their knowledge background. In fact, it is a technical issue. In EFL, the @ operator is only used to switch between agents, and we preserve this setting in our model. But EFL is a world-based logic, so it does not have the same problems that arise in a state-based account such like dealing with different information bases of varying cardinalities. If we set the shifting to occur between different states of different cardinalities, i.e., @ operator switches between information states that have different worlds and agents, then we may encounter various technical problems that need to be resolved, such as defining the difference between information states, determining whether the agent or cardinality is different, or whether both are different. These questions are left for future research.

Might-claims adhere to agents' perspectives, so in the context of a dialogue, distinct participants may validly affirm and contest a particular *might*-claim. Expressivist analysis for epistemic *might* sheds light on why Daiyu and Xiren rightly assert as they do, as each agent expresses her own attitude. There is no mystery as to why they stand in disagreement, since a single agent cannot hold beliefs that render both utterances appropriate.

However, if the disputed proposition is "For what I (Xiren) know, Baoyu might be in the Bamboo Lodge." the *might*-claim should be formalized as $@_x \blacklozenge p$. In this case, for Qingwen and Daiyu, what they need to judge is $@_x \blacklozenge p$ instead of plain *might*-formula $\blacklozenge p$. To judge the proposition from Daiyu or Qingwen's perspective, we need to add $@_d$ or $@_n$ in front of the formula $@_x \blacklozenge p$, respectively, namely $@_d @_x \blacklozenge p$ (for Daiyu, Baoyu might be in the Bamboo Lodge for Xiren) and $@_n @_x \blacklozenge p$ (for Qingwen, Baoyu might be in the Bamboo Lodge for Xiren). However, in our model, we have also confirmed the validity of the principles of hybrid logic and EFL:

$$@_i @_j \phi \leftrightarrow @_j \phi$$

Thus, it follows that regardless of the number of @ operators incorporated, the proposition $\blacklozenge p$ will ultimately be evaluated in reference to $s_a[g(x)]$. Consequently, it will be assessed as assertable by the model.

Agent	"Baoyu might be in the Bamboo Lodge for Xiren"	Judgement	Formalization
Xiren (x)	$@_x \blacklozenge p$	assertable	$\mathcal{M}, s_a \vDash @_x \blacklozenge p$
Qingwen (n)	$@_n @_x \blacklozenge p$	assertable	$\mathcal{M}, s_a \vDash @_n @_x \blacklozenge p$
Daiyu (d)	$@_d @_x \blacklozenge p$	assertable	$\mathcal{M}, s_a \vDash @_d @_x \blacklozenge p$

In order to explain the case of disagreement, especially to address the issue in what sense we can disagree with others, we employ the @ operator and locality effect as suggested by expressivist analysis for epistemic *might*. Specifically, our formalization allows us to check if a *might*-claim can be evaluated from the related agent's perspective, and if so, the agent can express agreement or disagreement with the claim. We achieve this by distinguishing between the translations of sentences such as "for what Xiren knows, Baoyu might be in the Bamboo Lodge" ($@_x \blacklozenge p$) and "Baoyu might be in the Bamboo Lodge" ($\blacklozenge p$).

In this section, we proposed a combined semantics to address the perspective-sensitivity of epistemic *might*. This goal is achieved through the combination of the locality of $\blacklozenge$ from BSEL and hybrid-style operators that enable perspective shifting.

It is noteworthy that the non-classical properties of epistemic *might*, such as epistemic contradictions and free choice effects, are preserved in our proposed semantics. This is due to the fact that the semantics continues to adhere to a state-based approach and its treatment of BSEL formulas aligns with the original BSEL model.

In the next section, we examine the *might*-claim in attitude reports and knowledge, and take into account knowledge and belief.

3.5 Whose knowledge is being expressed?

A predicate of personal taste is sensitive to an agent or group of agents whose tastes or preferences are relevant to its interpretation. The challenge lies in determining the relevant agent. Parallel behaviours between epistemic *might* and predicates of personal taste suggest that *might* is also sensitive to a knower or knowers. Epistemic *might* reports the knower's knowledge or epistemic state. Consequently, a similar question arises: whose knowledge is being expressed?

In this section, we first examine the role of the knower(s) in the interpretation of *might*-claims in belief reports, incorporating belief and informational modalities into our language. Then, we investigate the case of *might*-claims embedded under the verb *know*. To account for the factivity of knowledge, while also accommodating the perspective-sensitivity of epistemic *might*, we propose a BSEL-style definition of knowledge in a multi-agent model, where the factivity of knowledge is defined as presupposition rather than solely reflexivity.

3.5.1 Epistemic *might* under belief

As previously discussed in Section 2, the identification of *might*-claim asserters in the doxastic context is made clearer.

- In a believe-*might* sentence, like "Daiyu believes it might be raining", the knower to whom the embedded *might* clause is linked is Daiyu, the holder of the attitude.
- If a believe-*might* sentence embedded under another verbs of believe or think, such as "Daiyu believes that Baoyu believes that it might be raining", the knower is Baoyu who is the immediate subject of the *might*-claim.

In the analysis of the BSEL framework, the speaker's role is crucial in determining the interpretation of a sentence. However, there are instances where the speaker's knowledge or mental state does not impact the evaluation of a believe-*might* sentence. For instance, despite Daiyu not believing that it might be raining and having no evidence of rain, she can assert the sentence “Baoyu believes that it might be raining.” This is because the embedded *might*-claim pertains solely to Baoyu, and its evaluation requires only an examination of Baoyu’s beliefs. Furthermore, we distinguish the following two sentences in (11-a) and (11-b).

(11) Daiyu, as the speaker, asserts the following two sentences:

a. Baoyu believes that it might be raining.

b. I believe that Baoyu believes that it might be raining.

The assertion in (11-a) expresses Baoyu’s belief, whereas the assertion in (11-b) expresses Daiyu’s belief of Baoyu’s doxastic state. In the case of (11-a), the speaker’s perspective (Daiyu’s) does not influence the evaluation of the sentence, thereby she just describes Baoyu’s belief. However, in the case of (11-b), Daiyu’s perspective is implicated, even though the *might*-claim is only linked to Baoyu’s perspective. This is because in the sentence, Daiyu’s belief seems to generate Baoyu’s belief.

The BSEL framework cannot capture the subtle difference between (11-a) and (11-b). The two cases will be formalized in the same way. This is due to the fact that a BSEL model just represents the speaker’s perspective in its evaluation parameter. The perspectives of other agents are derived from the speaker’s information state, and the framework is unable to independently express the perspectives of other agents. As a result, if the speaker does not acknowledge the subject’s belief, or if the speaker doesn’t have the epistemic supervision to the subject, then subject’s belief of will be rejectable in the evaluation parameter s.

In order to differentiate between cases (11-a) and (11-b), we now consider the beliefs of agents in the multi-agent models. We extend the language $\mathcal{L}$ to include informational and doxastic modalities $\mathbf{I}$ and $\mathbf{B}$. The expressions $\mathbf{I}\phi$ and $\mathbf{B}\phi$ are respectively interpreted as “the agent has information of ϕ” and “the agent believes that ϕ”. The duals are defined in the standard manner, such that $\langle\mathbf{I}\rangle = \neg\mathbf{I}\neg$ and $\langle\mathbf{B}\rangle = \neg\mathbf{B}\neg$.

In accordance with the logic BSEL, the modality $\mathbf{I}$ is the standard $\mathbf{S}5$ modality, which is defined in terms of the equivalence relation that is already included in the multi-agent model $\mathcal{M}$. The epistemic modality $\mathbf{B}$ is represented by means of plausibility relations, hence it is necessary to incorporate a plausibility ordering $\leq_{a,w}$ into the model.

Now the multi-agent model is defined as: $\mathcal{M} = \langle W, A, R, \leq_{a,w}, g, V\rangle$, where $\leq_{a,w}$

is the plausibility ordering among possible worlds at w for agent $a \in A$. Intuitively, for every world v and v', $v \leq_{a,w} v'$ says that at the world w, v is at least as plausible as v' for the agent a. Then the semantic clauses for **I** and **B** can be defined.

$\mathcal{M}, s_a \vDash \mathbf{I}\phi$ iff $\forall i \in s_a : \mathcal{M}, R_{a_i}(w_i)[a_i] \vDash \phi$

$\mathcal{M}, s_a \dashv \mathbf{I}\phi$ iff $\forall i \in s_a$: there is a non-empty subset $t \subseteq R_{a_i}(w_i)[a_i]$, $t \dashv \phi$

$\mathcal{M}, s_a \vDash \mathbf{B}\phi$ iff $\forall i \in s_a$: $Dox_{a_i}(w_i)[a_i] \vDash \phi$

$\mathcal{M}, s_a \dashv \mathbf{B}\phi$ iff $\forall i \in s_a$: there is a non-empty subset $t \subseteq Dox_{a_i}(w_i)[a_i]$, $t \dashv \phi$

where the abbreviations are defined as follows:

- For a set Γ, $\Gamma[a_i] = \{< v, a_i > \mid v \in \Gamma\}$
- $R_a(w) = \{v \mid wR_a v\}$
- $R_{a_i}(w_i)[a_i] = \{< v, a_i > \mid v \in R_{a_i}(w_i)\}$
- $Min_{\leq_{a,w}}(R_a(w)) := \{v \in R_a(w) \mid \forall w' \in R_a(w) : w' \leq_{a,w} v \Rightarrow v \leq_{a,w} w'\}$
- $Dox_a(w) = Min_{\leq_{a,w}}(R_a(w))$
- $Dox_{a_i}(w_i)[a_i] = \{< v, a_i > \mid v \in Dox_{a_i}(w_i)\}$

The informational formula $\mathbf{I}\phi$ is supported in s_a if and only if for every $i \in s_a$, ϕ is supported in every equivalence class of w_i for agent a which is generated in term of relation R_a; $\mathbf{I}\phi$ is anti-supported in s_a if and only if for every equivalence class there exists a subset that anti-supports ϕ. A doxistic formula $\mathbf{B}\phi$ is supported in s_a if and only if ϕ is supported in every doxistic set of w_i for agent a; and it is anti-supported in s_a provided each doxistic set has a subset that anti-supports ϕ.

Associated with the operator $@_n$, we can capture the difference shown in (11). Let variable p be the proposition "it is raining," and nominal n and m indicate "Daiyu" and "Baoyu" respectively.

(12) a. "Baoyu believes that it might be raining" uttered by Daiyu.
b. Formalization: $@_n @_m \mathbf{B} \blacklozenge p$

(13) a. "I believe that Baoyu believes that it might be raining" uttered by Daiyu.
b. Formalization: $@_n \mathbf{B} @_m \mathbf{B} \blacklozenge p$

The formulas of this example are both assertable in the model depicted by Figure 3.3.

The depicted figure illustrates a simple model in which the points represent pairs of the world and agent, and the black line represents the equivalence relations[1]. The information

1 For this simple model, we do not distinguish the relations R_a and R_b, since R is an equivalence relation and all worlds are accessible from each other.

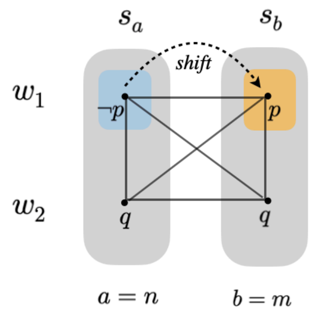


Figure 3.3 Superiority of what others believe

state s_a consists of $\langle w_1, a\rangle, \langle w_2, a\rangle$, while s_b consists of $\langle w_1, b\rangle, \langle w_2, b\rangle$. The value of proposition p is $V(p) = \langle w_1, b\rangle$. The blue and yellow areas in the figure depict the doxastic sets of agents a and b, respectively. In this model, the most plausible worlds for both agents a and b are the same, namely w_1. But $\langle w_1, a\rangle \notin V(p)$, while $\langle w_1, a\rangle \in V(p)$. It implies that, a doesn't believe p but b believes. Since there is a possible shifting from $\langle w_1, a\rangle$ to $\langle w_1, b\rangle$, the agent a can learn what b believes.

In this model, the interpretation of (12-b) leads to a move to the information state s_b, where the evaluation of $\blacklozenge p$ takes place. It is important to note that the operator $@_n$ does not play any role in this interpretation process.

To interpret the formula in (13-b), we proceed as follows: First, moving to Daiyu's perspective (s_a) by interpreting $@_n$. The interpretation of **B** operator allows us to obtain Daiyu's doxastic set $\{w_1\}$. Next, on w_1, shift the perspective from Daiyu's to Baoyu's by interpreting the $@_m$ operator, thereby jumping to Baoyu's perspective. Because w_1 also is the most plausible world for Baoyu. Then jump from Daiyu's doxastic set to Baoyu's doxastic set, and this move implies that Daiyu believes that Baiyu believes something. Finally the resulting point (the yellow area depicts) supports $\blacklozenge p$. So Daiyu believes that Baoyu believes $\blacklozenge p$ is assertable, even Daiyu herself doesn't believe $\blacklozenge p$.

As shown by the example, in both instances the *might*-claim is directly related to the subject Baoyu. However, there exits a scenario where the speaker does not believe what Baoyu believes, which is hard to formalize in BSEL models. We illustrate this by models depicted in the Figure 3.4 below.

In this model, the blue and yellow areas represent the doxastic sets of agents a and b, respectively. The proposition $@_n\mathbf{B}@_m\mathbf{B}\blacklozenge p$ is not assertable as the most plausible world for a is w_2, where q holds for both a and b. The shifting from $\langle w_2, a\rangle$ occurs only to the point $\langle w_2, b\rangle$, which lies outside the doxastic set of b. So it implies that the agent a

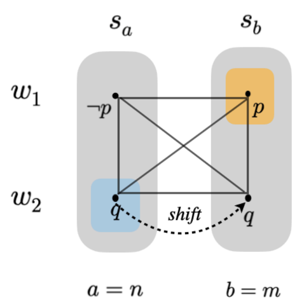


Figure 3.4 Believing differently

doesn't believe that the agent b holds a belief of p. However, contrary to BSEL models, we can still derive the assertability of $@_m\mathbf{B}\blacklozenge p$, namely the agent b believes p, since the most plausible world w_1 for b supports p. Therefore, if a denotes Daiyu and b denotes Baoyu in this model, then it can derive that the speaker, Daiyu, does not believe that Baoyu believes $\blacklozenge p$, but the statement "Baoyu believes that it might be raining." is still assertable.

Furthermore, the use of the binder operator $\downarrow n$ enables us to capture indexicals within the context of the verb *believe*. For example:

(14) a. I believes that Daiyu (m) believes that I am hungry (q).
b. Formalization: $\downarrow n\ \mathbf{B}@_m\mathbf{B}@_n q$

In the context of a given model $\mathcal{M}$ and information state s_a, the speaker is assumed to be represented by the current information state s_a. The operator $\downarrow n$ is used to bind the reference of n to the current agent in the information state s_a, thus enforcing n to refer to the speaker. The interpretation process begins with $\mathbf{B}$ from the perspective of the speaker (n), followed by a switch to Daiyu's perspective. Finally, the interpretation of q is performed with respect to the speaker's perspective, through the use of $@_n$, which takes into account the binding of n to the speaker via $\downarrow n$.

Note that the indexical "I" in sentence (14-a) must refer to the same agent–the speaker. Failure to use the operator "$\downarrow n$" in the formalization may result in the interpretation of the indexical "I" within the scope of Daiyu's belief referring to other agents intead of the speaker.

3.5.2 Knowledge

In this section, we proceed to formalize the concept of knowledge in multi-agent models. To begin, we consider the epistemic agents who possess knowledge of know-*might* sentences.

3.5.2.1 The knower

In Chapter 2, a comprehensive analysis was conducted on the distinction between "*know-might*" and "*believe-might*" sentences. One of the key differences highlighted was the fact that, in the interpretation of any "know-*might*" sentence, the speaker's perspective holds a significant influence. To reiterate, consider the following example: "Ronguofu might be in Hangzhou"[1] (in fact Ronguofu is in Jinling, not in Hangzhou).

(15) Ronguofu might be in Hangzhou.

a. Daiyu believes that Ronguofu might be in Hangzhou.

b. # Daiyu knows that Ronguofu might be in Hangzhou.

The sentence in (15-a) is acceptable due to the fact that the embedded *might*-claim is exclusively linked to the agent Daiyu's perspective. As the speaker, despite being aware that Ronguofu is not located in Hangzhou, you can still assert the sentence as Daiyu is entitled to hold an erroneous belief. Your perspective does not play a role in this assertion. Conversely, the sentence in (15-b) is deemed to be infelicitous. This is because such a sentence is dependent on both the perspectives of the speaker and the subject. Specifically, when uttering this know-sentence, you presuppose what Daiyu knows to be true from your perspective.

In BSEL, we can account for this by defining factivity of knowledge as presupposition. However, as mentioned, it cannot account for the following sentence:

(16) Daiyu knows that Baoyu knows that it might be raining.

It is noteworthy to distinguish this case from the scenario involving a "*believe-might*" sentence in multiple embeddings. In (16), the *might*-claim is not solely tied to the knowledge of the immediate subject, but rather linked to three agents: the speaker, Daiyu, and Baoyu.

In conclusion, for *know-might* sentences, the embedded *might*-claims are consistently linked with the speaker's perspective and the perspective of the subject of each *know* verb in the sentence.

1 This example is from a famous Chinese novel *A Dream of Red Mansions*. In the novel, Rongguofu is located in Jinling, not in Hangzhou.

3.5.2.2 Factivity

In BSEL, knowledge is defined as the formula $(\mathbf{I}\phi)_\phi$, where $\mathbf{I}$ represents the informational modality. The formula ϕ_ψ is introduced as a formula of presupposition and is interpreted as "ϕ presupposes ψ". This definition of knowledge allows for the incorporation of the speaker's perspective in its evaluation, as discussed in Chapter 2.

In order to establish a definition of knowledge within multi-agent models, it is necessary to first define the formula of presupposition.

$$\mathcal{M}, s_a \vDash \phi_\psi \quad \text{iff} \quad \mathcal{M}, s_a \vDash \phi \text{ and } \mathcal{M}, s_a \vDash \psi$$
$$\mathcal{M}, s_a \mathrel{=\!\!\mid} \phi_\psi \quad \text{iff} \quad \mathcal{M}, s_a \mathrel{=\!\!\mid} \phi \text{ and } \mathcal{M}, s_a \vDash \psi$$

ϕ_ψ is supported in s_a if and only if ϕ is supported and also its presupposition is supported; and ϕ_ψ is anti-supported if and only if ϕ is anti-supported but its presupposition is still supported[1].

However, defining knowledge as $(\mathbf{I}\phi)_\phi$ presents a problem. For instance, consider the sentence "Daiyu (m) knows that it might be raining", which is formalized as $@_m K \blacklozenge p$. Based on our previously defined semantics, the distribution of the operator $@_n$ applies to presuppositions, such that $@_n(\phi_\psi) \vDash (@_n\phi)@_n\psi$. Hence, we have $@_m K \blacklozenge p = @_m((\mathbf{I}\blacklozenge p)\blacklozenge p)$, which implies that $(@_m \mathbf{I}\blacklozenge p)@_m\blacklozenge p$. The formulas $\mathbf{I}\blacklozenge p$ and $\blacklozenge p$ are evaluated with respect to the information state $s_{g(d)}$, meaning only one agent's perspective (Daiyu's) is considered in the evaluation of the sentence, and the speaker's perspective, represented by the current information state s_a, has not been involved.

Another limitation of this definition is that it only enables the derivation of factivity of knowledge in restricted situations. If knowledge is defined as $(\mathbf{I}\phi)_\phi$, the following inferences cannot be obtained.

- $@_n K\phi \nvDash \phi$, e.g. Daiyu knows $\phi \not\Rightarrow \phi$
- $K @_n K\phi \nvDash \phi$, e.g. I know that Daiyu knows $\phi \not\Rightarrow \phi$
- $@_n K @_m K\phi \nvDash @_n\phi$, e.g. Daiyu knows that Baoyu knows that $\phi \not\Rightarrow \phi$.

The operator $\downarrow n$ enables the factivity reasoning in situations where the proposition in question is known by any arbitrary agent.

- $K \downarrow m @_m K\phi \vDash \phi$

This says that "If I know that anyone knows ϕ, then ϕ holds for me."

These problems motivate us to propose a revised definition of knowledge in the multi-

1 This is because negation of an expression does not change its presuppositions.

agent model in the following way[1]:

$$\mathbf{K}_n\phi := (@_n\mathbf{I}\phi)_\phi$$

We read the formula $K_n\phi$ as "the agent named n knows ϕ".

$$\mathcal{M}, s_a \vDash \mathbf{K}_n\phi \quad \text{iff} \quad \mathcal{M}, s_a \vDash @_n\mathbf{I}\phi \text{ and } \mathcal{M}, s_a \vDash \phi$$

$$\mathcal{M}, s_a \dashv \mathbf{K}_n\phi \quad \text{iff} \quad \mathcal{M}, s_a \dashv @_n\mathbf{I}\phi \text{ and } \mathcal{M}, s_a \vDash \phi$$

Based on this semantics, it can be inferred that

$$K_n\phi \vDash \phi$$

Based on this definition, ϕ is interpreted related to the current information state s_a representing the perspective of the speaker. Intuitively, "n knows ϕ" means n has information of ϕ and ϕ is presupposed by the speaker.

Furthermore, the placement of a binder $\downarrow n$ in front of $K_n\phi$ expresses the indexical proposition "I (the agent with the index n) know that ϕ".

In this section, we extended our multi-agent framework by incorporating elements of knowledge and belief. Our framework acknowledges that, in most instances, the knower of the *might*-claim in belief reports is its immediate subject. However, if the indexical referring to the speaker is presented in the sentence, the speaker's perspective becomes a factor in the interpretation of the *might*-claim. These considerations are incorporated in our expanded multi-agent model, which takes into account belief.

We have proposed a refined definition of knowledge in a multi-agent system. The essence of this definition is the recognition that the speaker's perspective plays a crucial role in the interpretation of knowledge statements. By means of this definition, we are able to derive factivity of knowledge.

In what follows, a comprehensive overview of the multi-agent model will be presented, and the potential predictions that can be derived from it will be concluded.

3.6 A multi-agent extension of BSEL: put it all together

3.6.1 The multi-agent setting

Definition 21 (Language $\mathcal{L}_{Multi}$)**:** *The language is based on three types of atoms: propositional variables $p \in Prop$, nominols $n \in ANom$, and the non-emptiness atom* NE.

$$\varphi ::= p \mid n \mid \texttt{NE} \mid \neg\varphi \mid \varphi \wedge \varphi \mid \varphi \vee \varphi \mid \blacklozenge\varphi \mid \mathbf{I}\varphi \mid \mathbf{B}\varphi \mid \varphi_\varphi \mid @_n\varphi \mid \ \downarrow n\ \varphi$$

The language $\mathcal{L}_{Ag}$ results from adding nominals and the hybrid-style operators $@_n$

1 We appreciate Jeremy Seligman's suggestions on this problem.

and $\downarrow n$ into the language of BSEL. $\blacklozenge$ stands for epistemic *might*. The operators $\mathbf{I}$ and $\mathbf{B}$ stand for informational modality and doxastic modality respectively. The satisfaction operator $@_n$ borrowed from EFL is for shifting perspectives, and we read $@_n\phi$ as "for agent named n, ϕ holds". The operator $\downarrow n$ behaves like a binder. The duals of I and B are defined as usual: $\langle\mathbf{I}\rangle = \neg\mathbf{I}\neg$, $\langle\mathbf{B}\rangle = \neg\mathbf{B}\neg$. Knowledge as a derived notion is defined as:

$$\mathbf{K}_n\phi := (@_n\mathbf{I}\phi)_\phi$$

Models for $\mathcal{L}_{Ag}$ are Kripke-style models of the form $\mathcal{M} = \langle W, A, R_a, \leq_{a,w}, g, V\rangle$.

- W is the set of possible worlds and A is the set of agents.
- R_a represents the indistinguishability relation for agent $a \in A$.
- $\leq_{a,w}$ is the plausibility ordering among possible worlds at w for agent $a \in A$.
- g is an assignment function: ANom$\rightarrow A$.
- V is an evaluation function from Prop to the subsets of $W \times A$, with $(w, a) \in V(p)$, representing p holds for a at w.

An information state s is defined as a set of indices i. An index i is a pair $i = \langle w_i, a_i\rangle$. We assume that for all $i, j \in s$, $a_i = a_j$. In fact there is only one agent in s. Hence we use the agent a in s as a subscript s_a to mark who the agent in this information state is.

Formulas of $\mathcal{L}_{Ag}$ are interpreted in $\mathcal{M}$ with respect to information states. The semantic clauses are defined as follows.

Definition 22 (Semantic clauses of $\mathcal{M}$)**:**

$\mathcal{M}, s_a \vDash p$	*iff*	$\forall i \in s_a : (w_i, a_i) \in V(p)$
$\mathcal{M}, s_a \Dashv p$	*iff*	$\forall i \in s_a : (w_i, a_i) \notin V(p)$
$\mathcal{M}, s_a \vDash n$	*iff*	$\forall i \in s_a,\ g(n) = a$
$\mathcal{M}, s_a \Dashv n$	*iff*	$\forall i \in s_a,\ g(n) \neq a$
$\mathcal{M}, s_a \vDash \mathrm{NE}$	*iff*	$s_a \neq \emptyset$
$\mathcal{M}, s_a \Dashv \mathrm{NE}$	*iff*	$s_a = \emptyset$
$\mathcal{M}, s_a \vDash \neg\phi$	*iff*	$\mathcal{M}, s_a \Dashv \phi$
$\mathcal{M}, s_a \Dashv \neg\phi$	*iff*	$\mathcal{M}, s_a \vDash \phi$
$\mathcal{M}, s_a \vDash \phi \vee \psi$	*iff*	$\exists t, t' : t \cup t' = s_a$ *and* $\mathcal{M}, t \vDash \phi$ *and* $\mathcal{M}, t' \vDash \psi$
$\mathcal{M}, s_a \Dashv \phi \vee \psi$	*iff*	$\mathcal{M}, s_a \Dashv \phi$ *and* $\mathcal{M}, s_a \Dashv \psi$
$\mathcal{M}, s_a \vDash \phi \wedge \psi$	*iff*	$\mathcal{M}, s_a \vDash \phi$ *and* $\mathcal{M}, s_a \vDash \psi$
$\mathcal{M}, s_a \Dashv \phi \wedge \psi$	*iff*	$\exists t, t' : t \cup t' = s_a$ *and* $\mathcal{M}, t \Dashv \phi$ *and* $\mathcal{M}, t' \Dashv \psi$
$\mathcal{M}, s_a \vDash @_n\phi$	*iff*	$\mathcal{M}, s_a[g(n)] \vDash \phi$
$\mathcal{M}, s_a \Dashv @_n\phi$	*iff*	$\mathcal{M}, s_a[g(n)] \Dashv \phi$

$\mathcal{M}, s_a \vDash \downarrow n\ \phi$ *iff* $\mathcal{M}^n_a, s_a \vDash \phi$

$\mathcal{M}, s_a \dashv \downarrow n\ \phi$ *iff* $\mathcal{M}^n_a, s_a \dashv \phi$

$\mathcal{M}, s_a \vDash \blacklozenge\phi$ *iff* *there is a non-empty subset* s'_a *of* s_a*, such that* $s'_a \vDash \phi$

$\mathcal{M}, s_a \dashv \blacklozenge\phi$ *iff* $\mathcal{M}, s_a \dashv \phi$

$\mathcal{M}, s_a \vDash \boldsymbol{I}\phi$ *iff* $\forall i \in s_a : \mathcal{M}, R_{a_i}(w_i)[a_i] \vDash \phi$

$\mathcal{M}, s_a \dashv \boldsymbol{I}\phi$ *iff* $\forall i \in s_a$: *there is a non-empty subset* $t \subseteq R_{a_i}(w_i)[a_i]$, $t \dashv \phi$

$\mathcal{M}, s_a \vDash \boldsymbol{B}\phi$ *iff* $\forall i \in s_a$: $Dox_{a_i}(w_i)[a_i] \vDash \phi$

$\mathcal{M}, s_a \dashv \boldsymbol{B}\phi$ *iff* $\forall i \in s_a$: *there is a non-empty subset* $t \subseteq Dox_{a_i}(w_i)[a_i]$, $t \dashv \phi$

$\mathcal{M}, s_a \vDash \phi_\psi$ *iff* $\mathcal{M}, s_a \vDash \phi$ *and* $\mathcal{M}, s_a \vDash \psi$

$\mathcal{M}, s_a \dashv \phi_\psi$ *iff* $\mathcal{M}, s_a \dashv \phi$ *and* $\mathcal{M}, s_a \vDash \psi$

where

- $i[g(n)] = \langle w_i, g[n]\rangle$
- $s_a[g(n)] = \{i[g(n)] \mid i \in s\}$
- $\mathcal{M}^n_a = \langle W, A, \sim_a, \leq_{a,w}, g^n_a, V\rangle$, *where* g^n_a *is a n-variant of g.*
- *For a set* Γ, $\Gamma[a_i] = \{< v, a_i > \mid v \in \Gamma\}$
- $R_a(w) = \{v \mid wR_a v\}$
- $R_{a_i}(w_i)[a_i] = \{< v, a_i > \mid v \in R_{a_i}(w_i)\}$
- $Min_{\leq_{a,w}}(R_a(w)) := \{v \in R_a(w) \mid \forall w' \in R_a(w) : w' \leq_{a,w} v \Rightarrow v \leq_{a,w} w'\}$
- $Dox(w) = Min_{\leq_{a,w}}(R_a(w))$
- $Dox_{a_i}(w_i)[a_i] = \{< v, a_i > \mid v \in Dox_{a_i}(w_i)\}$

Definition 23 (Logical consequence)**:**

$\phi \vDash \psi$ *iff for all* $\mathcal{M}, s_a$: $\mathcal{M}, s_a \vDash \phi \Rightarrow \mathcal{M}, s_a \vDash \psi$

3.6.2 Some results

The multi-agent framework lies in the combination of BSEL and EFL models. By combing these two models, the result model can have the outcomes or predictions that are generated by each individual model. However these results will not be presented as they can be deduced effortlessly based on the established definition of the model. Instead, we present the results only obtained from the combination.

3.6.2.1 Different perspectives for knowledge

In Section 4, we give an example to show that the same *might*-claim can be evaluated differently from different perspectives. Furthermore, this phenomenon can also be observed with respect to *know*-sentences, as previously argued a *know*-sentence is sensitive to the speaker. A *know*-sentence can be assertable only if the speaker agrees with the content of this sentence. For example, consider the scenarios depicted in Table 3.1, and evaluate the sentence S ="Daiyu(m) knows that it is raining(p)".

Speaker	Context	Judgement	Formalization
Daiyu (m)	Daiyu knows p	assertable	$s[g(m)] \vDash K_m p$
Baoyu (n)	Baoyu knows that Daiyu knows p (so Baoyu knows p)	assertable	$s[g(n)] \vDash K_m p$
Baoyu (j)	Baoyu doesn't know p	not assertable	$s[g(j)] \nvDash K_m p$
Xuepan (k)	Xuepan knows p but doesn't know that Daiyu knows p	rejectable	$s[g(k)] \Dashv K_m p$

Table 3.1 Different perspectives for know-sentences

As the speaker, Daiyu is expressing her own knowledge. Baoyu knows p and hence he agrees with Daiyu on p. So the sentence S is assertable for both of them in the sense that the verb *know* can be used in the sentence for them to describe Daiyu's mental (epistemic) state. However, since Baoyu lacks knowledge of p, he would rather use another verb (like *believe* or *being certain* etc), namely a verb which would not commit Baoyu to the truth of p. As a result, the sentence S is not assertable for Baoyu. On the other hand, Xuepan knows p, yet he is not aware of Daiyu's knowledge of p. So Xuepan rejects the sentence S. The results of evaluation are shown in Table 3.1.

Notice that Baoyu's judgment for S is possible in our semantics, which is not assertable or rejectable. Because our multi-agent model is a state-based system that can reject bivalence (Aloni (2022)).

It is worth stressing the difference between Baoyu's judgement and Xuepan's. We define knowledge $\mathbf{K}_n\phi$ as a formula that triggers presupposition $(@_n\mathbf{I}\phi)_\phi$. The semantic clauses for $\mathbf{K}_n\phi$ consist of the assertability condition of $@_n\mathbf{I}\phi$ as well as the condition that ϕ is presupposed. So there are two ways of denying a $\mathbf{K}$-formula: (1) deny the informational formula $@_n\mathbf{I}\phi$; (2) deny its presupposition ϕ. For Xuepan and Baoyu, both of them deny the statement S. However S is rejectable for Xuepan, but not rejectable or assertable for Baoyu. Because Xuepan denies the informational part but validates the presupposition. While Baoyu denies the presupposition, so neither support nor anti-support condition of S holds for him. In other words, it is predicted to be a presupposition failure.

3.6.2.2 Factivity of knowledge

In multi-agent models, the dimension of agency has resulted in a more nuanced understanding of the factivity of knowledge. The factivity of knowledge now indicates that a given agent knows ϕ if and only if ϕ is assertable for that agent. Similarly, the complement of the verb *know* in a *know*-sentence uttered by a given agent is assertable for that speaker. We list the inferences of factivity in what follows.

Fact 10: *$K_a\phi \vDash \phi$, in particular $K_a\blacklozenge\phi \vDash \blacklozenge\phi$.*

Fact 10 can be proved by the definition of knowledge.

Fact 11:

- $K_nK_m\phi \vDash \phi$.
- $K_nK_m\phi \vDash K_n\phi$
- $K_nK_m\phi \vDash K_m\phi$

The inferences in Fact 11 show that ϕ is assertable for each knower.

Fact 12: *$K_n\phi \nvDash @_n\phi$, except for the case where $s_a = s_a[g(n)]$.*

One can find a counter-example similar to the example in Chapter 2 that shows $[\mathbf{I}]\blacklozenge p \nvDash \blacklozenge p$, where $[\mathbf{I}]$ is informational modality in BSEL.

3.7 Summary

In this chapter, we present an extension of the BSEL model by incorporating a multi-agent framework. This integration of the two-dimensional semantics of EFL and BSEL results in a more comprehensive representation of agency in epistemic statements.

The proposed model characterize the aspects of agency in *might*-claims, belief reports, and *know*-sentences, allowing for the identification and representation of the relevant agent or knower. Furthermore, the multi-agent model addresses issues related to agency in epistemic modals, such as faultless disagreement.

As future research, we plan to extend our understanding of knowledge in this semantics by investigating concepts such as common knowledge and distributed knowledge.

In Chapters 2 and 3, our focus lies on the exploration of epistemic *might*, knowledge, and belief. The puzzle arising from the WEM reflects the a property that bridges semantic and pragmatic reasoning, prompting us to consider both its semantic attributes and pragmatic behaviors when capturing a modality, akin to our definition of knowledge that combines with presupposition. In the forthcoming chapters, our attention will shift to the examination of modal expressions of desire. We will commence by conducting a systematic analysis of the WEM, addressing questions such as whether it is semantic or pragmatic, and whether the puzzles generating from it pose challenges to a monotonic semantics. Following a thorough examination of monotonicity puzzles in the context of *want* we will address to these inquiries before delving into the investigation of the semantics of *want*.

CHAPTER 4 THE OVERTONE OF MONOTONICITY UNDER DESIRE

In the present chapter, our primary objective is to address the framework question through a systematic examination of the WEM. As a case study, we will explore a series of puzzles in the context of desire, as the tension exhibited by monotonicity puzzles involving desire is both more intuitive and widely discussed.

A notable departure from the previous two chapters is our adoption of first-order language for our discussion. This decision stems from our belief that the replacement of predicates in monotonic inference gives rise to additional inferences. Consequently, a meticulous analysis of predicates is necessary to comprehend this phenomenon. And thus, we will employ the first-order version of BSML as the foundational framework for formal account.

4.1 Introduction

It is controversial whether the attitude verbs expressing desire (e.g.*want*, *hope*) license monotonic inferences in their scope. Several examples have been proposed as evidence for a non-monotonic behaviour under such verbs which led to the development of non-monotonic semantics for bouletic modalities (see Heim (1992); Lassiter (2011); Levinson (2003); Xie and Jialiang (2022), etc.). Firstly, we present three typical puzzles challenging a monotonic semantics for *want*.

4.1.1 Asher's puzzle

The first example is Asher's puzzle, which demonstrates how upward entailments in desire environments can lead to problems. Heim reports the puzzle in Heim (1992), in which she replaced *hope* in the original case with *want*. This puzzle involves the scenario where Nicholas wants a free trip on the Concorde, but is not willing to pay the 3,000 dollars he believes it would cost. Despite the fact that that taking a free trip on the Concorde, of course, implies taking a trip on the Concorde, it is still true that he wants a free trip, but false that he wants a trip on the Concorde.

(1) a. Taking a free trip $\Rightarrow$ taking a trip

b. Nicholas wants a free trip on the Concorde.

c. Nicholas wants a trip on the Concorde.

A monotonic semantics for ”want” would require that if Nicholas wants a free trip on the Concorde, he must also want a trip on the Concorde. However, this is clearly not the case, as Nicholas is not willing to pay the cost of the trip.

4.1.2 Heim’s example

In the same article Heim also presents another puzzle about desire, which has a similar logical form of the Good Samaritan paradox[1] in deontic logic. The scenario involves a speaker who says they want to teach on Tuesdays next semester. However, this does not necessarily imply that they want to teach next semester.

(2) a. Teaching on Tuesdays next semester ⇒ Teaching next semester

b. I want to teach on Tuesdays next semester.

c. I want to teach next semester. (Heim (1992))

This paradox arises because it is entirely possible to utter the sentence ”I want to teach on Tuesdays next semester” in a situation where the speaker doesn’t want to teach at all. This example highlights the difficulty in making inferences about desire, as the speaker’s true intentions may not be captured by their words.

4.1.3 Ross’ paradox under desire

Ross’ Paradox (Ross (1944)) is a well-known paradox from deontic logic[2] that concerns the monotonic inference ‘Send the letter! ⇒ Send the letter or burn it!’, which is logically valid but intuitively invalid. In Crnič (2011), Crnič observed that embedding the two non-finite clauses under *want* leads to equally paradoxical results. The scenario involves John, who wants to send the letter. However, the conclusion licenses the inference that John wants to send the letter or burn it.

(3) a. John wants to send the letter.

b. John wants to send the letter or burn it. (Crnič (2011))

1 The paradox was first introduced by Prior (Prior (1958)). ⊚*A*qvist (Aqvist (1967)) reproduced a similar but more popular version of it and provided a clear formulation.

2 See McNamara and Van De Putte (2022) which gives a good overview.

This paradox arises because the conclusion suggests that John is okay with burning the letter, which may not be the case.

There are many other puzzles challenging monotonic semantics of *want*, e.g. the examples in Lassiter (2011) and Crnič (2011). In this chapter, we focus on the typical examples that have the inferential mode of upward entailment. Assuming the monotonic semantics of *want*, the inferential patterns of the above puzzles are summarized in Table 4.1.

	Inferential modes
Asher's puzzle	$A \to B$, $[\mathtt{Want}]A \to [\mathtt{Want}]B$
Heim's puzzle	$(A \wedge B) \to A$, $[\mathtt{Want}](A \wedge B) \to [\mathtt{Want}]A$
Ross' paradox	$A \to (A \vee B)$, $[\mathtt{Want}]A \to [\mathtt{Want}](A \vee B)$

Table 4.1 Inferential modes of the three puzzles

As shown above, Asher's puzzle includes an entailment from a notion to a more general one (a free trip ⇒ a trip)[1]. The inference $A \to (A \vee B)$ in Ross' paradox is classically valid, following the rule of disjunction introduction in classical logic.

Notice that Heim's example is not totally analogous to Asher's puzzle, although they appear to share some similarity. The logical form of Heim's example is considered to be closer to the Good Samaritan paradox in which there is an entailment from a conjunction to one of its conjunct. This is because the proposition "Teaching on Tuesday" has a finite phrase "on Tuesday" that limits the teaching to Tuesday, i.e. there is a teaching and it happens on Tuesday. And hence it will be formalized to be a conjunction $A \wedge B$ instead of an atomic proposition. The premise (2-b) is true so the conjunction $A \wedge B$ can occur in the scope of *want*, however, A that follows $A \wedge B$ cannot appear under *want* alone.

In conclusion, these three examples demonstrate the complex and often paradoxical nature of desire inferences. They highlight the need for a more nuanced approach to understanding the behavior of verbs expressing desire and the difficulties in making inferences based on such verbs. The development of a unified account for these puzzles, could help to provide a more comprehensive understanding of the nature of desire and its implications for language and reasoning. As a result, we expect to establish further evidence that the puzzles don't provide counter-examples against interpreting bouletic modals monotonically.

The chapter is structured as follows: We first examine two prominent existing analyse

1 The generalization of the predicate can also be viewed as a case of conjunction elimination: $\mathtt{TRIP}(x) \wedge \mathtt{FREE}(x) \Rightarrow \mathtt{TRIP}(x)$.

in Section 2. Then we will propose that all three puzzles can be reduced to the cases of Free Choice (FC) inferences in Section 3. The process leading to paradox will be clarified by assigning disjunctive meaning to predicates, especially in the case with no overt disjunctive statements. In Section 4, we first review Aloni's proposal QBSML (Aloni and van Ormondt (2023)), and based on it we add a reinterpretation function to account for disjunctive meaning of predicates. We discuss deontic cases and other applications in Section 5. Finally we conclude in Section 6.

4.2 Existing analyses of the puzzles

In this section, we will delve into the details of the two prominent existing analyses of desire verbs such as *want*. First, we will review the semantics for *want* proposed by Heim (Heim (1992)) and then examine the semantics proposed by Von Fintel (Von Fintel (1999)). We will compare and contrast the two proposals and examine their solutions to the puzzles presented in Introduction. This review of the existing literature will provide a foundation for our discussion of a unified account for the puzzles on upward monotonicity under desire in the next section.

4.2.1 Heim's preference semantics of *want*

Heim's theory in Heim (1992) offers a new perspective on the meaning of *want*. One of her key contributions is the idea that *want* operates as an operator over belief. Heim argues that *want* modifies the doxastic state of an agent, ranking the complement of *want* as more desirable than its negation. The purpose of Heim's analysis of *want* in relation to belief was to address Karttunen's generalization (Karttunen (1974)), which states that the presuppositions in the complement of *want* are believed by the agent α. Another significant contribution is the concept of *want* having a semantics of preference. This ranking of preferences informs the speaker's beliefs about what they would like to be the case, and this information is used in the interpretation of *want* in natural language.

Heim first provide a static version. She adopts Stalnaker's explanation (Stalnaker (1968)), and argues that wanting something can be defined as preferring it to certain relevant alternatives, which are the possibilities that the agent believes will be realized if he does not get what he wants. For example, John wants to eat seafood for dinner means that John thinks that if there is seafood for dinner, he will be in a more desirable world than if there is not.

To give a formal definition, Heim employs a version of the semantics proposed by Lewis (Lewis (1973)) for counterfactual conditionals and Stalnaker (Stalnaker (1968)) for conditionals in general, and defines the meaning of *want* as follows:

- Heim's static semantics for *want* (Heim (1992))

  "α wants ϕ" is true in w iff for every $w' \in Dox_\alpha(w)$:
  every ϕ-world maximally similar to w' is more desirable to α in w than any non-ϕ-world maximally similar to w'.

Heim highlights that one advantage of this analysis over Hintikka's simpler treatment of bouletic modals in term of an accessibility relation (Hintikka (1969))[1], is that the new analysis predicts the failure of upward monotonicity under desire. It means that if A entails B, α wants A does not necessarily entail α wants B. For example, based on this definition Asher's puzzle is solved by blocking the monotonic inference. Nicholas's doxastic-worlds where he flies for 3,000 dollars are worse off than in minimally differing worlds where he doesn't fly at all. But in this case the premise could also be true. Since those doxastic-worlds where he does get a free trip are better than similar worlds where he doesn't or he flies and pays. This predication is based on the preference ordering for Nicholas: free trip > no trip > \$3000 trip.

According to Hintikka's analysis, the expression of desire "I want to teach on Tuesdays next semester" is predicted to be false, as the accessible worlds that are compatible with everything the speaker desires do not include a requirement to teach at all. On the other hand, Heim's analysis predicts that this desire sentence is true, while the statement "I want to teach next semester" is falsified. This is because in all doxastic worlds, the speaker either teaches on Tuesdays or teaches the same load on different weekdays. And the former is considered more desirable.

In Heim's proposal, the similarity between worlds can be defined in term of a similar function Sim_w:

- $Sim_w(p) = \{w' \in W : w' \in p$ and w' resembles w no less than any other world in $p\}$

This function Sim_w takes a proposition p and a world w as input and finds all the worlds that are maximally similar to w with respect to the proposition p. Similarity is determined by membership in propositions, but Heim does not specify what these propositions are. The function operates twice in the semantics of *want*, once to generate a set of

1 A bouletic relation is defined as $R_{bl}(w)$: $\{v \in W \mid v$ conforms to what the agent wants in $w\}$.

maximally similar ϕ-worlds to a doxastic world w, and another to generate a set of non-ϕ worlds that are maximally similar to w. However, this leads to a potential flaw. Because the comparison of desirability is not always between doxastic worlds, which undermines to fully explain Karttunen's generalization that Heim intends to formalize. To overcome this, Heim suggests a second version in which the comparison of worlds is limited to the α's doxastic worlds.

- Heim's dynamic semantics for *want* (Heim (1992))

  $c + \alpha$ wants $\varphi = \{w \in c :$
  $\forall w' \in \mathrm{Dox}_\alpha(w), \mathrm{Sim}_{w'}(\mathrm{Dox}_\alpha(w) + \varphi) <_{\alpha,w} \mathrm{Sim}_{w'}(\mathrm{Dox}_\alpha(w) + \neg\varphi)\}$

Based on this definition, if an agent α wants ϕ, then a doxastic world with a proposition ϕ being true is more desirable than a similar doxastic world where ϕ is false. This semantics allows the truth of a *want*-sentence to be determined by comparing the desirability of **doxastic** worlds, which are determined by the subject's beliefs. And desirability of non-belief worlds does not affect the truth of a want-report.

Notice that the second version is defined in the form of **context change potential** (CCP, see Heim (1983)) semantics, which is a dynamic treatment. This move is also needed for her goal of formalizing Karttunen's generalization on presupposition projection behavior. Heim's analysis suggests that the presuppositions of embedded clauses under the verb *want* are not projected beyond the *want*-clause. As per Karttunen's generalization, she proposed that the presuppositions of the complement of a *want*-sentence are satisfied only if the attitude holder believes them. Consequently, Heim rejects Hintikka's approach to defining *want* as inadequate to account for belief. Heim proposed a definition that incorporates the attitude holder's beliefs, given that presuppositions are projected.

Therefore, based on the second semantics, the sentence "I want to teach next semester" would be predicted to be a presupposition failure instead of falsity, so that the *want*-sentence is assigned no truth value. Since I believe that I will teach next semester, there exists no non-teaching world in my doxastic alternatives. Consequently, the function sim_w would not be able to identify any not-teaching-worlds as the most similar ones to compare with a teaching-belief world.

4.2.2 Von Fintel's monotonic semantics

Von Fintel in Von Fintel (1999) proposed a refinement of a Hintikka-style semantics for *want*, which is now sometimes referred to as a "best worlds" semantics. According to von

Fintel, a desire ascription claims that the best worlds according to the agent's preferences are all worlds in which the prejacent is true. In other words, among the agent's doxastic worlds, the ones that are best by the lights of the agent's preferences are all prejacent-worlds. This analysis is formulated as follows:

- Von Fintel's semantics for *want* (Von Fintel (1999))

  $[\![\text{want}]\!] = \lambda p.\lambda\alpha.\lambda w : p \nsubseteq dox_{\alpha,w}$ and $p \cap dox_{\alpha,w} \neq \varnothing^1$. $\forall w' \in best_{\alpha,w}(dox_{\alpha,w}) : p(w') = 1$

where p is the prejacent of the desire ascription, α is the agent, w is the world of evaluation, $dox_{\alpha,w}$ is the doxastic set of α in w, and $best_{\alpha,w}$ is the set of worlds that are best by the lights of α's preferences in w.

In this analysis, two additional ingredients are introduced to account for the fact that many of our desires are "realistic" and to explain why the first ingredient doesn't lead to us wanting anything that we believe to be true. The first addition is the constraint that the worlds up for comparison are constituted by the agent's doxastic set (similar to Heim's semantics). So von Fintel's semantics, also predicts Heim's example "I want to teach" to have a presupposition failure (because $p \cap dox_{\alpha,w} = \varnothing$). The second addition is the requirement that the prejacent needs to not already be settled by the agent's doxastic set. This is there to explain why the first addition doesn't lead to us wanting anything that we believe to be true.

Von Fintel's analysis is argued to be Strawson-monotone (Von Fintel (2018)), as we argued by examples (4) and (5) in Section 1. He wants to account for the intuition that the perceived infelicity of monotonic inferences may be attributable to pragmatic influences.

Based on von Fintel's semantics, he considers examples of contradictory statements in pragmatics, such as the example of Asher's puzzle, which is semantically valid but appears to be pragmatically infelicitous. He argues that the apparent contradiction can be resolved by considering the information introduced in the discourse and the order in which the statements are made. Consider a discourse where a ticket salesman asks Nicholas:

(4)
a. Salesman: Do you want a trip on the Concorde?
b. Nicholas: No, I don't want a trip on the Concorde.
c. Salesman: Do you want a free trip on the Concore?
d. Nicholas: Yes, I do want a free trip on the Concorde.

1 This condition serves for presupposition. If $p \cap dox_{\alpha,w} = \varnothing$ there will be a presupposition failure.

In the discourse, the saleman first ask Nicholas about his desire for a trip on the Concorde in (4-a) and then ask him about a free trip in (4-c). This maneuver results in a change in Nicholas's beliefs, and he expresses a desire for a free trip in response to the second question. To evaluate the statements in (4-b) and (4-d) in this discourse accurately, they should be assessed with respect to distinct modal bases, namely different doxastic alternatives. This is because the second question in (4-c) introduces new information that alters Nicholas's mental state. Thus, statement in (4-b) doesn't contradict with (4-d), since they are envaluated with respect to different contexts. More details of this pragmatic analysis can be found in Von Fintel (1999) and Crnič (2011).

Notice that, in von Fintel's analysis, an enlargement is introduced by the premise. This differs from our interpretation under WEM theory. In our account, an enlargement also occurs, but it is triggered by monotonicity rather than the premise. We will show this in what follows.

4.2.3 A need for a pragmatic account

Table 4.2 summarizes the predictions of Heim's and Von Fintel's analyses concerning the problems posed by the puzzles discussed in Introduction.

	Ross' Paradox	**Asher's Puzzle**	**Heim's example**
Von Fintel	Semantically valid	Pragmatically invalid	Presupposition failure
Heim	Semantically valid	Semantically invalid	Presupposition failure

Table 4.2 Comparing Predictions of Existing Analyses

Heim argues that upward monotonicity fails in the presence of desire verbs due to the interaction between the presuppositions of the desire verb and the entailments of the desire context. While von Fintel's analysis differs from Heim's in that it seeks to explain the puzzles by appealing to the interaction between the desire verb and pragmatic effects.

Notice that neither Von Fintel nor Heim has a ready explanation for the puzzling inferences in (3). So they both predict Ross's paradoxical inference to be valid.

In the following section we are going to propose a pragmatic explanation of Ross' paradox and extend such an explanation to the cases of Asher's Puzzle and Heim's example. A pragmatic account of the three puzzles would allow us to maintain a monotonic semantics for bouletic verbs. Evidence in favor of a monotonic semantics comes from examples like the following:

(5) I don't want to eat any fast food.

As Von Fintel noted (Von Fintel (2012)), according to the standard theory of Negative Polarity Items (NPIs), a downward-entailing environment is required for NPI *any* to be licensed. So if *want* is not upward monotonic, then *not want* will not produce a downward entailing environment. Furthermore, if we refuse the upward monotonicity of desire, we would fail to account for the intuitive infelicity of the following sentence:

(6) Jones wants to buy a green sweater, but she doesn't want to buy a sweater. Zimmermann (2006)

In addition, as noted in Haslinger (2020), research in psychological reasoning furnishes supplementary, showing that individuals are always influenced by pragmatic effects when requested to evaluate entailment (refer to Stenning and van Lambalgen (2008); Counihan (2008)).

Therefore we will explore if a upward monotonic analysis of desire can be maintained by combing it with a pragmatic account which can handle these puzzles uniformly. In the next section, we will propose that all three puzzles can be reduced to the cases of Free Choice (`FC`) inferences.

4.3 Reinterpretation of predicates

4.3.1 Free choice inferences under desire

Consider again the problem posed by Ross' paradox, which is repeated below:

(7) a. John wants to send the letter.
   b. John wants to send the letter or burn it.

The sentence in (7-b) may convey that John has a positive attitude towards both sending and burning the letter. This positive attitude shows that the two options concerning the letter are acceptable to John, i.e., at least they are not unwanted. In Crnič (2011), Crnič uses *it is ok...* to represent this weaker attitude. We will employ the same expression in our proposal. It follows that the inference in (8) is licensed:

(8) a. John wants to send the letter or burn it.

b. $\rightsquigarrow$ It is ok for John to send the letter, **and** it is ok for him to burn it.[1]

As Crnič observed the inference in (8) has the same form of well known FC inferences. One of the observations in the FC literature is that a conjunctive meaning of possibility modals can be derived from a disjunctive necessity modal statement. It can be denoted as $\Box$-free choice, which is different from the commonly held FC inferences from $\Diamond$-modal to $\Diamond$-modal ($\Diamond$-FC).

(9) a. **$\Box$-free choice**
John ought to mail the letter or burn it. So he is allowed to mail it and he is allowed to burn it. $[\Box(A \vee B) \rightsquigarrow \Diamond A \wedge \Diamond B]$

b. **$\Diamond$-free choice**
You may take chocolate or ice cream. So you may take chocolate, and you may take ice cream. $[\Diamond(A \vee B) \rightsquigarrow \Diamond A \wedge \Diamond B]$

In the chapter, we analyse the semantics of desire verbs in Hintikka's style. In what follows, the modality *want* is treated as $\Box$, and the weaker desire *it is ok...* is interpreted as $\Diamond$. Then we can apply the principle in (9-a) to the inference in (8) as follows:

(10) $\Box_J$ (SEND $\vee$ BURN) $\rightsquigarrow \Diamond_J$SEND $\wedge \Diamond_J scalebox 0.8[1]$BURN

As shown, the conclusion of a weaker desire to burn the letter can be drawn as a FC inference. However, it is absurd to deduce from someone's willingness to send a letter her acceptance of burning it. This is why the puzzle appears to be paradoxical.

As mentioned in Chapter 1, the core proposal is that a weakening effect, triggered by monotonicity under attitudes, is the principal cause of the phenomena and puzzles under investigation which are labeled as the *weakening effect triggered by monotonicity*. We fundamentally assume that the weakening resulting from applications of monotonicity gives rise to pragmatic effects, similar to the free choice and ignorance inferences triggered by disjunctive statements. The weakening could result in the audience interpreting the conclusions drawn from monotonicity by the bigger models, thereby leading to a possible pragmatic failure.

A similar strategy to the puzzle can be found in Crnič (2011), and some literature also applies the principle of FC to analyze the deontic Ross' paradox concerning imperatives

1 The inference is drawn when we read the sentence (7-b) in a way where the scope of the modality *want* is over the disjunctive statement, namely it has a wide scope.

or permissions, e.g. Von Fintel (2012); Aloni (2007). In the present chapter, we will provide a formal account of FC inferences under bouletic modalities based on Aloni (2022), where the FC inference is assumed to be derived as a product of the interaction of literal meanings and pragmatic factors. Before that, we explore if the analysis proposed above is also applicable to the other two puzzles.

4.3.2 Free choice triggered by non-disjunctive statements: Asher's puzzle

Asher's puzzle is repeated below:

(11) a. Nicholas wants a free trip on the Concorde.
b. Nicholas wants a trip on the Concorde.

The most apparent difference between Ross' paradox and Asher's puzzle is that there is no overt disjunctive statement in the latter example, which would be required to trigger a FC inference. Our strategy doesn't seem to work for this case. But the puzzle hints that there is some additional information that can be acquired from the conclusion (11-b), which leads us to refute the inference. So what makes the Asher's puzzle paradoxical, we will argue, is very similar to what causes Ross' paradox.

What gives us to think that the reasoning in (11) is odd? Our proposal is that a crucial factor is the additional information which is conveyed by the conclusion (11-b). Accordingly we are lead to believe that he is ok with a free trip but also with a trip which is not free. However the premise only states Nicholas' desire for a free trip, and it provides nothing to justify his being ok with a non-free trip. Therefore it seems that the following inference can be drawn from (11-b):

(12) Nicholas is ok with a non-free trip on the Concorde.

This would be the inference we would trigger if the original sentence were reinterpreted as a disjunction:

(13) Nicholas wants a trip on the Concorde $\Leftrightarrow$ Nicholas wants a free or non-free trip on the Concorde.

In what follows we will argue for the feasibility for such reinterpretation. Imagine a scenario where a kid, who really wants chocolate, is shopping with her mother. In this situation, what she cares about is "will my mom buy chocolate for me?". This can be seen as a *Question under Discussion* (QUD, see Roberts (1996)) in this context. The answers to this polar question are introduced as topics which bring our attention to the predicate *chocolate*, which hence becomes salient. Furthermore, we propose that salient predicates can influence the way people interpret the predicates that semantically include them. Suppose the mother says to the girl "Let's go to buy some snacks", then the girl will wonder "Are we going to buy chocolate?". In this case, the predicate *snacks* is naturally reinterpreted via *chocolate* as a disjunctive predicate "chocolate or any other snacks that are not chocolate". Namely, snacks is semantically divided into two categories: one includes chocolates, and the other includes snacks which are not chocolates.

We propose that a predication, in conversations, can be reinterpreted as a disjunctive statement and so convey a disjunctive meaning. In Asher's case, the predicate *trip* in (1-c) can be interpreted as *free trip or non-free trip* since the being free of the trip is made salient by the premise. This reinterpretation is possible because $\forall x Qx$ and $\forall x((Qx \wedge Px) \vee (Qx \wedge \neg Px))$ are logically equivalent and therefore can be substituted salva veritate. We need to emphasize that, in our proposal, a predicate only can be disjunctively reinterpreted by predicates that describe the objects within its denotation, i.e. by predicates which are semantically included ($P \subset Q$). We call them *sub-predicates*. This analysis can be supported by **Exemplar Theory** in cognitive psychology.

Exemplar theory is a theoretical framework in cognitive psychology that explains how people categorize and recognize objects and events (see Brooks (1987); Thibaut and Gelaes (2006); Rosch (1978); Nosofsky (1986), etc.). The theory argues that people form mental representations, or prototypes, of objects and events based on their experiences. These prototypes are then used to recognize and categorize new stimuli.

This approach is based on the idea that memory is comprised of specific experiences or examples, rather than abstract, generalized categories. When an individual encounters a new stimulus, they compare it to the stored exemplars in their memory. If the new stimulus is similar enough to a stored exemplar, it is categorized based on that exemplar. If the new stimulus is not similar to any stored exemplars, a new exemplar is created in memory.

The central claim of exemplar theory is that people use their stored exemplars to make judgments about the categorization and recognition of new stimuli. This approach is in contrast to *prototype theories*, which hold that people have abstract, generalized prototypes in their memory that they use to categorize new stimuli.

One of the key contributions of exemplar theory is its ability to account for variations in categorization and recognition. For example, exemplar theory can explain why people categorize objects based on their context, or why people can recognize objects even if they are different from their stored exemplars. In conclusion, exemplar theory is a valuable tool for understanding how people process and organize information in their memory.

Evidence from the exemplar theory shows that concepts are typically represented as remembered (hence salient) instances. So in a conversation, although the semantics of a predicate (such as *trip*) is clear to the recipients, people typically focus only on a part of the objects in its extension. As a result we propose a reinterpretation theory that a disjunctive meaning of a predicate can be obtained by a reinterpretation, consisting of the union of the salient objects (first disjunct) and the remaining objects (second disjunct). The latter is normally denoted as the negation of the former.

So far, we have proposed that predicates in conversations can be reinterpreted leading to disjunctive representations. In addition, we have two constraints:

- In a disjunctive reinterpretation of a predicate Q, the disjuncts are always *sub-predicates* of Q.
- The sub-predicates should be *salient* in the context.

In this way, the process from *Nicholas wants a trip* to *Nicholas is ok with a free trip and is ok with a non-free trip* can be clarified as follows: (14-a) undergoes a disjunctive reinterpretation and then the disjunctive statement gives rise to a $\texttt{FC}$ inference.

(14) a. Nicholas wants a trip.
b. $\Leftrightarrow$ Nicholas wants a free trip or a non-free trip.
[by disjunctive reinterpretation]
c. $\rightsquigarrow$ Nicholas is ok with a free trip and he is ok with a non-free trip.
[by $\Box$-$\texttt{FC}$ principle]

As a result, Nicholas' attitude of accepting a non-free trip is derived, which is unwarranted by the premise. Consequently, the monotonic reasoning in (11) appears to fail.

It is noteworthy that the reinterpretation theory we proposed in this thesis shows the potential for a wide range of applications. For instance, in the work of Ju (2023), the theory is used to examine the issue of presupposition projection for factive predicates in both English and Mandarin Chinese languages.

4.3.3 Free choice triggered by non-disjunctive statements: Heim's example

Heim's example is repeated in the following:

(15) a. I want to teach on Tuesday next semester.
     b. I want to teach next semester.

What the agent wants are teaching events that are scheduled for Tuesdays next semester. The sentence (15-a) is the kind of desire expression that is used every day. However (15-b) is not the intended meaning conveyed by (15-a) but rather a logical consequence derived from the premise. Why does (15-b) seem inappropriate? Similarly to the previous case, we propose that (15-b) can be represented as a disjunctive statement which generates a free choice inference showing the agent's attitude towards accepting teaching events on any day, which is unjustified:

(16) a. I want to teach next semester.
     b. I want to teach on Tuesday or teach on the other days next semester.
     c. I'm ok to teach on the days that are not Tuesday next semester.

As in the case of Asher's puzzle we argue that (16-c) is pragmatically drawn from (16-a). To further clarify the analysis for example (16), it is important to address any potential doubts about its semantic validity through the use of monotonicity. If someone argues that the conclusion (15-b) is false because it implies that the agent has no desire to teach at all, then the premise about the desire to teach on Tuesday should also be considered false. However, this may only be the case if the original meaning of the premise was that if the agent is required to teach next semester (she has to do it), she will only accept to teach on Tuesdays. In other words, the example demonstrates a conditional desire. Thus, our evaluation of the conclusion (15-b) should also be based on the same condition, meaning it should be rephrased under the condition "if I have to teach":

(17) If I have to teach next semester,
     a. I'm ok to teach on Tuesday next semester.
     b. I'm ok to teach next semester.

In such case, our analysis is still applicable, as the predicate *teach* can be reinterpreted as *teach on Tuesdays or teach on the non-Tuesdays*. Then we can apply the principle of $\Diamond$-FC in (9-b):

(18) a. I'm ok to teach next semester.

b. I'm ok to teach on Tuesdays or the other days next semester.

c. I'm ok to teach on Tuesdays next semester **and** I'm ok to teach on other days next semester.

Notice that we started by saying that Heim's example has similar logical structure with the Good Samaritan paradox. Because in the premise (15-a), the proposition in the scope of desire can be formulated as a conjunction: "having teachings next semester **and** the teachings are scheduled on Tuesdays", but not both of these two conjuncts are desired. In fact, there are also more direct Good Samaritan cases under desire. For instance,

(19) a. The school wants that students who struggle with math attend tutoring.

b. The school wants that students struggle with math.

We do not want to claim that our analysis of disjunction applies to all the Good Samaritan cases under desire. We believe that other solutions also could be given. For example, we take a general analysis to the Good Samaritan paradox from deontic logic (e.g Tomberlin and Mcguinness (1977)). The main cause of the paradox in (19) is the scope ambiguity. Let us assume that no teacher wants students struggle with math. The relative clause "who struggle with math" is an expression that serves as an adjective so it can behave as the predication. So we have the following equivalence:

(20) Students who struggle with math attend tutoring. $\Leftrightarrow$ Students struggle with math and these students attend tutoring.

Two readings of (19-a) can be observed:

(21) **Wide scope**:

The school wants that (students struggle with math **and** these students attend tutoring)

$\Box(\exists x(\mathtt{MATH}x \wedge \mathtt{TUTO}x))$

(22) **Narrow scope**:

There are some students struggle with math and the school wants them to attend tutoring

$\exists x(\mathtt{MATH}x \wedge \Box\mathtt{TUTO}x)$

If we take the wide scope reading of (19-a), then (19-a) will be false because of the assumption ($\Box\neg\exists x\mathtt{MATH}x$). If we take the narrow scope reading, then (19-b) cannot be derived by monotonicity.

Similarly, for the example in (15), we could also thinking of taking the proposition 'having teachings next semester' outside the scope of desire to block the monotonic inference. This is also an effective solution. But neither the analysis of disjunction we proposed nor the solution of scope ambiguity justify a non monotonic semantics for the desire modality.

In conclusion, we have shown the functionality of the WEM theory in resolving the presented puzzles, particularly in conjunction with the application of disjunctive reinterpretation. In the theory, we use the $\mathtt{FC}$ to clarify why the three puzzles appear to be paradoxical. In the next section, we will employ a logical framework *Bilateral State-based Modal Logic* (BSML) and its first-order version (QBSML) based on Aloni (2022); Aloni and van Ormondt (2023) to give a formal account for the $\mathtt{FC}$ inferences.

4.4 A formal account for free choice

4.4.1 Quantified bilateral state-based modal logic (QBSML)

In Aloni (2022), Aloni proposed a formal account of $\mathtt{FC}$ inference in Bilateral State-based Modal Logic (BSML). On this account $\mathtt{FC}$ and related inferences are a consequence of a tendency operative in conversation which Aloni calls *neglect zero*. On the neglect-zero hypothesis, people when interpreting a sentence construct representations of the world and in doing so they systematically neglect models that validate sentences by virtue of some empty configuration (zero-model). We have introduced the propositional BSML in previous chapter. In what follows we introduce the first-order version QBSML (Aloni and van Ormondt (2023)) and apply it to our analysis with some additional assumptions.

The core idea of BSML is providing a method to formalize the *neglect-zero* effects by employing the non-emptiness atom ($\mathtt{NE}$) from team logic (Yang and Väänänen (2017)), which is added as a part of the syntax.

Definition 24 (Language $\mathcal{L}_{\mathcal{D}}$)**:** *The language $\mathcal{L}_{\mathcal{D}}$ is defined as following in BNF:*

Term $t ::= c \mid x$

Formula $\phi ::= P^n t_1 ... t_n \mid \neg\phi \mid \phi \wedge \phi \mid \phi \vee \phi \mid \exists x \phi \mid \Box \phi \mid$ NE

$P^n \in \mathbf{P}^n$ ($n \in \mathbb{N}$) stand for predicate constants. $c \in \mathcal{C}$ and $x \in \mathcal{V}$ stand for individual constants and variables respectively. The quantifier $\exists x$ is similar to the existential quantifier in first order logic. The operator $\Box$ stands for the verb *want*, and $\Diamond$, which stands for *it is ok...*, can be defined as the dual of $\Box$: $\Diamond\phi := \neg \Box \neg\phi$. Notice that we don't dive into the semantics of *want* in this chapter. We only focus the pragmatic account for monotonic failure under *want*. So we treat $\Box$ as a bouletic modality in a classical Hintikka's style, which behaves like a quantifer over possibile worlds and is interpreted in term of a bouletic relation.

A *pragmatic enrichment function* (denoted by $[\cdot]^+$) is defined in terms of a systematic intrusion of NE in the process of interpretation:

Definition 25 (Pragmatic enrichment [Aloni and van Ormondt (2023)): *] A pragmatic enrichment function is a mapping $[\cdot]^+$ from the NE-free fragment of language $\mathcal{L}_\mathcal{D}$ to $\mathcal{L}_\mathcal{D}$ such that:*

- $[Pt_1...t_n]^+ = Pt_1...t_n \wedge NE$
- $[\neg\phi]^+ = \neg[\phi]^+ \wedge NE$
- $[\phi \vee \psi]^+ = ([\phi]^+ \vee [\psi]^+) \wedge NE$
- $[\phi \wedge \psi]^+ = ([\phi]^+ \wedge [\psi]^+) \wedge NE$
- $[\Box\phi]^+ = \Box[\phi]^+ \wedge NE$
- $[\exists x\phi]^+ = \exists x[\phi]^+ \wedge NE$

Definition 26 (Model $\mathcal{M}$)**:** *A model for $\mathcal{L}_\mathcal{D}$ is a tuple $\mathcal{M} =< W, D, R, I >$:*

- *W is a set of possible worlds.*
- *D is a non-empty set which is assumed to be constant across worlds.*
- *R is bouletic accessibility relation on W. $R(w)$: $\{v \in W \mid$ where wRv means that v conforms to what the agent wants in $w\}$. We define the following abbreviation:*

  $R(w) = \{v \in W \mid wRv\}$.

- *I: $W \times \mathcal{C} \cup \boldsymbol{P}^n \to D \cup \mathcal{P}(D^n)$[1] is an interpretation function which assigns entities to individual constants and sets of n-tuples of entities to predicate letters relative to worlds $w \in W$:*

1 $\mathcal{P}(S)$ indicates the power set of S.

$$I(w)(\gamma) = \begin{cases} d \in D & if\ \gamma \in \mathcal{C} \\ S^n \subseteq D^n & if\ \gamma \in \boldsymbol{P}^n \end{cases}$$

An interpretation of a term in a world can be defined as following:

Definition 27 (Interpretation of a term t)**:**

$$[\![t]\!]_{\mathcal{M},i} = \begin{cases} g_i(t) & if\ t \in \mathcal{V} \\ I(w_i)(t) & if\ t \in \mathcal{C} \end{cases}$$

QBSML, as a quantification extension of BSML, is based on state-based semantics. This means that the formulas in QBSML are interpreted with respect to information states. However, in a first order modal model, an information state is not merely a set of possible worlds. Variable assignments must also play a role in its definition.

An information state in the first-order modal framework is defined as a set of indices. An index i is a pair i=$< w_i, g_i >$ where $w_i \in W$ and $g_i = \mathcal{V} \to D$. By the indices, an information state can encode information about the value of variables in worlds. Now we can define the semantic clauses for atoms and connectives.

Definition 28 (Semantics[Aloni and van Ormondt (2023))**:** *] Let $\mathcal{M}$ be a model for language $\mathcal{L}_{\mathcal{D}}$ and s be a state, we define what it means for a formula ϕ to be supported or anti-supported at s.*

$$\begin{array}{lll}
\mathcal{M}, s \vDash P^n t_1 ... t_n & \textit{iff} & \forall i \in s : \langle [\![t_1]\!]_{\mathcal{M},i}, ..., [\![t_n]\!]_{\mathcal{M},i} \rangle \in I(w_i)(P^n) \\
\mathcal{M}, s \dashv P^n t_1 ... t_n & \textit{iff} & \forall i \in s : \langle [\![t_1]\!]_{\mathcal{M},i}, ..., [\![t_n]\!]_{\mathcal{M},i} \rangle \notin I(w_i)(P^n) \\
\mathcal{M}, s \vDash \neg\phi & \textit{iff} & \mathcal{M}, s \dashv \phi \\
\mathcal{M}, s \dashv \neg\phi & \textit{iff} & \mathcal{M}, s \vDash \phi \\
\mathcal{M}, s \vDash \phi \vee \psi & \textit{iff} & \exists t, t' : t \cup t' = s \textit{ and } \mathcal{M}, t \vDash \phi \textit{ and } \mathcal{M}, t' \vDash \psi \\
\mathcal{M}, s \dashv \phi \vee \psi & \textit{iff} & \mathcal{M}, s \dashv \phi \textit{ and } \mathcal{M}, s \dashv \psi \\
\mathcal{M}, s \vDash \phi \wedge \psi & \textit{iff} & \mathcal{M}, s \vDash \phi \textit{ and } \mathcal{M}, s \vDash \psi \\
\mathcal{M}, s \dashv \phi \wedge \psi & \textit{iff} & \exists t, t' : t \cup t' = s \textit{ and } \mathcal{M}, t \dashv \phi \textit{ and } \mathcal{M}, t' \dashv \psi \\
\mathcal{M}, s \vDash \mathrm{NE} & \textit{iff} & s \neq \emptyset \\
\mathcal{M}, s \dashv \mathrm{NE} & \textit{iff} & s = \emptyset
\end{array}$$

A simple model, depicted in Figure 4.1, is used to illustrate the semantics.

QBSML models differ from BSML models. In this example, there are two worlds designated by letters w, along with a subscript indicating the fact in that world. For example, w_{Pa} represents the world w where a is P. A dot in the model represents an index that is a pair of a world and an assignment. The information state s in this model consists of the three indices: $s = \{\langle w_{Pa}, g[x/a] \rangle, \langle w_{PaPb}, g[x/a] \rangle, \langle w_{PaPb}, g[x/b] \rangle\}$. $g[x/d]$ represents

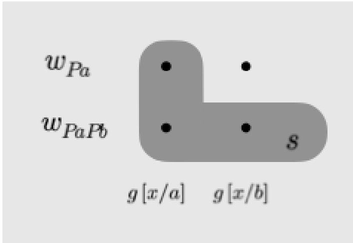


Figure 4.1 QBSML model

the assignment function g that maps the variable x onto an object $d \in D$.

In this model, we can derive that $s \vDash Px$ since in each $i \in s$, $g(x) \in I(w)(P)$. It also implies that $s \vDash [Pa \vee Pb]^{+}$, as there are two non-empty subset of indices which support Pa and Pb respectively.

We now examine the semantics for the quantifier and modal operator. To define this, we first present three ways of extending a state.

Definition 29:

a. $g[x/d] := (g \backslash \{\langle x, g(x)\rangle\}) \cup \{\langle x, d\rangle\}$

b. $i[x/d] := \langle w_i, g_i[x/d]\rangle$.

If the variable $x \notin \mathrm{dom}(g)$, then it can be added and set to the value d as $g[x/d]$. An index $i[x/d]$ is the result of adding x assigned to d to dom(g). Notice that if x is already in dom(g), then the operation $g[x/d]$ just indicate resetting the value of x.

Based on these notions, we can define three different ways of x-extension of a state s (see Dekker (1993); Aloni (2001)):

Definition 30 (x-Extensions of a state s)**:** • *Individual x-extension of s.*

$s[x/d]$:=$\{i[x/d] \mid i \in s\}$, *for some* $d \in D$.

- *Universal x-extension of s.*

  $s[x]$:=$\{i[x/d] \mid i \in s$, *and* $d \in D\}$.

- *Functional x-extension of s.*

  $s[x/h]$:=$\{i[x/d] \mid i \in s$, *and* $d \in h(i)\}$, *for some function* $h: s \to \mathcal{P}(D) \backslash \varnothing$.

The individual and universal x-extension of s, are the states which result by extending s with the assignment $g[x/d]$ for some $d \in D$ and $g[x/d]$ for all $d \in D$ respectively. And from the definition, it follows that the individual and universal extensions are examples of functional extensions in which any state t where for each index $i \in s$ there exists an index $j \in t$ such that $j = i[x/d]$ for some $d \in D$.

Now we can define the semantic clauses for quantification and modality.

Definition 31 (Quantifier and model operator[Aloni and van Ormondt (2023)):

$$\mathcal{M}, s \vDash \exists x\phi \quad \textit{iff} \quad \mathcal{M}, s[x/h] \vDash \phi, \textit{ for some function } h: \ s \to \mathcal{P}(D) \backslash \varnothing$$
$$\mathcal{M}, s \dashv \exists x\phi \quad \textit{iff} \quad \mathcal{M}, s[x] \dashv \phi$$
$$\mathcal{M}, s \vDash \Box\phi \quad \textit{iff} \quad \forall i \in s,\ R(w_i)[g_i] \vDash \phi$$
$$\mathcal{M}, s \dashv \Box\phi \quad \textit{iff} \quad \forall i \in s, \textit{ there is } X \subseteq R(w_i) \textit{ and } X \neq \varnothing \textit{ and } X[g_i] \dashv \phi$$

The abbreviation $X[g_i]$ is defined as follows:

- $X[g_i] = \{< w, g_i > \mid w \in X\}$

The quantifier in QBSML is defined in a manner consistent with the conventions established in team logic (Kontinen and Väänänen (2009)) and dynamic semantics. Since we define $\Diamond\phi$ as $\neg\Box\neg\phi$, the semantics for $\Diamond\phi$ can be given as:

$$\mathcal{M}, s \vDash \Diamond\phi \quad \text{iff} \quad \forall i \in s, \text{there is } X \subseteq R(w_i) \text{ and } X \neq \varnothing \text{ and } X[g_i] \vDash \phi$$
$$\mathcal{M}, s \dashv \Diamond\phi \quad \text{iff} \quad \forall i \in s,\ R(w_i)[g_i] \dashv \phi$$

Desire modals are interpreted in a Hintikka's style, which means that $[\texttt{Want}]\phi$ is evaluated in term of an accessibility relation. Given that states in this framework are comprised of sets of world-assignment pairs, so we interpret $\Box\phi$ or $\Diamond\phi$ by evaluating φ with respect to a state constructed by combining the worlds that are accessible from w_i with g_i for each relevant i.

4.4.2 Some results in QBSML

In this section, we show some results in QBSML, and argue the reason why we adopt it as the baseline framework to formalize our analysis.

In the framework, $\Box$- and $\Diamond$-FC inferences can be derived from pragmatically enriched disjunctions. Proofs can be found in Aloni and van Ormondt (2023), and we present the proof of $\Box$-free choice exclusively in this chapter.

Fact 13 (FC inference):

- $\Box$-*FC:* $[\Box(Pa \vee Pb)]^+ \vDash \Diamond Pa \wedge \Diamond Pb$
- $\Diamond$-*FC:* $[\Diamond(Pa \vee Pb)]^+ \vDash \Diamond Pa \wedge \Diamond Pb$

Proof: Suppose $\mathcal{M}, s \vDash [\Box(Pa \vee Pb)]^+$. It follows that $\forall i \in s$, $R(w_i)[g_i] \vDash [(Pa \vee Pb)]^{+1}$. Thus there must be non-empty t, t' such that $t \cup t' = R(w_i)[g_i]$ and $\mathcal{M}, t \vDash Pa$ and $\mathcal{M}, t' \vDash Pb$. Which means for every $i \in s$ there are non-empty subsets X, X' of $R(w_i)$

1 There are no free variable we can ignore the assignment component of the desire states.

and hence a non-empty sub-state $X[g_i] \vDash Pa$ and a non-empty sub-state $X'[g_i] \vDash Pb$. Therefore $\mathcal{M}, s \vDash \Diamond Pa \wedge \Diamond Pb$. $\square$

Notice that $\square$-free choice inferences can be easily derived via negations of universal alternatives in neo-Gricean approaches:

- $\square(\phi \vee \psi) + \neg \square \phi + \neg \square \psi \vDash \Diamond\phi \wedge \Diamond\psi$

However, some experiments show that distributive inferences may obtain without plain negated universal inferences (see Crnič et al. (2015)). Consider the following example:

(23) John wants to buy a Ferrari or a Porsche. $\rightsquigarrow$ John is ok with a Ferrari and he is ok with a Porsche.

In this situation, the conclusion in (23) does not seem to rely on John not wanting a Ferrari, and not wanting a Porsche.

(24) $\square(\phi \vee \psi) \rightsquigarrow \Diamond\phi \wedge \Diamond\psi$, even in absence of $\neg \square \phi$ and $\neg \square \psi$

Furthermore, the standard Gricean reasoning struggles to account for $\Diamond$-free choice inferences.

(25) $\Diamond(\phi \vee \psi) + \neg\Diamond\phi + \neg\Diamond\psi \nvDash \Diamond\phi \wedge \Diamond\psi$

(Q)BSML can capture these inferences, as well as other types of free choice inferences without relying on negated universal inferences. Furthermore, contrary to other semantic accounts of free choice also captures their cancellation under negation. As mentioned in Chapter 2, in (Q)BSML, free choice and ignorance inferences are taken to be pragmatic effects but not of the conversational implicature kind. Rather they follow from something else that speakers do in conversation, namely the neglect zero effects. This approach has received some preliminary confirmation by recent experiments (see Aloni (2022) for details).

Since our pragmatic account for the puzzles involves $\square$- and $\Diamond$-FC inferences, which can generate some unwarranted inferences. (Q)BSML is a suitable framework for formalizing our proposal due to its ability to formally derive these inferences.

4.4.3 Formalization of reinterpretation

In addition to free choice inferences, predicates also play an important role in our analysis. We have two constraints stated in Sec. 4.3. To address our puzzles, we need to formalise them in QBSML.

As argued, we do not think that a predicate can always be interpreted as a disjunctive statement. It happens only if one of its sub-predicates becomes **salient** in the context. To capture this, we define a *reinterpretation function* which only applies in case saliency is satisfied. Let `Nfo` be the set of all NE-free formulas of $\mathcal{L}_{\mathcal{D}}$, and `Prt` be the set of all atomic predication. A reinterpretation function $\|\texttt{Nfo}\|_{\texttt{Px}}$ is a mapping from $\texttt{Nfo}\times\texttt{Prt}$ to NE-free formulas of $\mathcal{L}_{\mathcal{D}}$:

Definition 32 (Reinterpretation function)**:**

- $\|Q\vec{x}\|_{P\vec{x}} = \begin{cases} P_Q\vec{x} \vee \neg P_Q\vec{x} & if\ P \subset Q \\ Q\vec{x} & otherwise \end{cases}$
- $\|\neg\phi\|_{P\vec{x}} = \neg \|\phi\|_{P\vec{x}}$
- $\|\phi \wedge \psi\|_{P\vec{x}} = \|\phi\|_{P\vec{x}} \wedge \|\psi\|_{P\vec{x}}$
- $\|\phi \vee \psi\|_{P\vec{x}} = \|\phi\|_{P\vec{x}} \vee \|\psi\|_{P\vec{x}}$
- $\|\exists x\phi\|_{P\vec{x}} = \exists x \|\phi\|_{P\vec{x}}$
- $\|\Box\phi\|_{P\vec{x}} = \Box \|\phi\|_{P\vec{x}}$

The function intuitively says that, a predicate Q can be reinterpreted syntactically as $\lambda x[P_Qx \vee \neg P_Qx]$ where P_Qx stands for $Px \wedge Qx$. The abbreviation is not a recursive definition since we expect $\neg P_Qx$ to stand for $\neg Px \wedge Qx$ instead of $\neg(Px \wedge Qx)$. If P is a sub-predicate of Q, namely if the denotation of P is the proper subset of the denotation of Q. Notice that, in (Q)BSML, classically logically equivalent formulas can give rise to different pragmatic effects under the pragmatic enrichment $[\cdot]^+$, e.g. it is possible to have a counterexample where $[\exists xQx]^+$ is supported but $[\exists x((Px \vee \neg Px) \wedge Qx)]^+$ is not. So it is not trivial to reinterpret a predicate disjunctively in our system.

Consequently, a formula $\varphi(\|Q\vec{x}\|_{P\vec{x}})$ can be rewritten as $\varphi(P_Q\vec{x} \vee \neg P_Q\vec{x})$ by replacing Q in φ by $\lambda x[P_Qx \vee \neg P_Qx]$. Then, with the function of pragmatic enrichment, the examples in which there is no overt disjunctive statements now can be explained by `FC`.

Now, let us revisit the puzzles in question. The Ross' paradox can be explained directly by referring to Fact 13.

(26) **Ross' paradox**

a. John wants to send the letter. $\Box\mathtt{SEND}a$

b. John wants to send the letter or burn it. $\Box(\mathtt{SEND}a \vee \mathtt{BURN}a)$

c. $\rightsquigarrow$ It is ok for John to send the letter. $\Diamond\mathtt{SEND}a$

d. $\rightsquigarrow$ It is ok for John to burn letter. $\Diamond\mathtt{BURN}a$

The constant variable a represents the letter. As shown in Aloni and van Ormondt (2023), the NE-free fragment of QBSML can be reduced to classical quantified modal logic. As with BSML, free choice principles are only applicable to pragmatically enriched formulas in QBSML. We call semantically valid the inferences validated by the NE-free fragment of QBSML and pragmatically valid the inferences holding for pragmatically enriched formulas. As depicted in Figure 4.2, the formula $\Box(\mathtt{SEND}a \vee \mathtt{BURN}a)$ is semantically derived from the premise $\Box\mathtt{SEND}a$ by monotonicity of $\Box$, but the inference from the premise to the pragmatically enriched formula $\Box[(\mathtt{SEND}a \vee \mathtt{BURN}a)]^+$ is blocked which hence is not assertable. However, according to $\Box$-FC in QBSML, $\Diamond\mathtt{SEND}a$ and $\Diamond\mathtt{BURN}a$ are drawn from $\Box[(\mathtt{SEND}a \vee \mathtt{BURN}a)]^+$ rather than the plain formula $\Box(\mathtt{SEND}a \vee \mathtt{BURN}a)$. Therefore we can safely use the monotonicity and free choice principle.

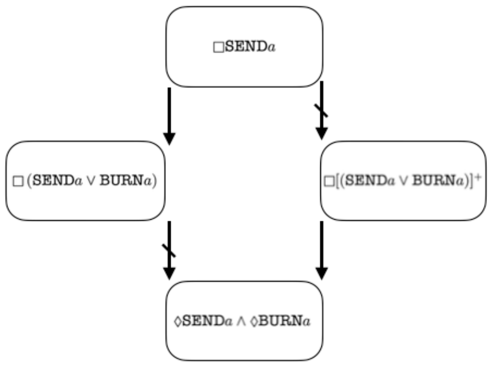


Figure 4.2 Solution to Ross' paradox

The resolution strategy for Asher's puzzle is akin to that of the Ross paradox. However, it crucially relies on the disjunctive reinterpretation of "TRIP".

(27) **Asher's puzzle**

a. Nicholas wants a free trip on the Concorde. $\Box\exists x\mathtt{FREE}\ x$

b. Nicholas wants a trip on the Concorde. $\Box\exists x\mathtt{TRIP}\ x$

c. It is ok for Nicholas to have a free trip. $\Diamond\exists x\mathtt{FREE}\ x$

d. It is ok for Nicholas to have a non-free trip. $\Diamond\exists x\neg\mathtt{FREE}\ x$

As depicted in Figure 4.3, the formula $\Box\exists x \mathtt{TRIP}\ x$ is semantically (monotonically) derived from the premise $\Box\exists x \mathtt{FREE}\ x$. But the pragmatically enriched formula containing the reinterpretation of $\mathtt{TRIP}$ as $\Box\exists x[\|\mathtt{TRIP}\ x\|_{\mathtt{FREE}}]^{+}$ is not justified by the premise and is, therefore, not assertable. Consequently, the attitude towards accepting a non-free trip is not semantically determined since it is only licensed by $\Box\exists x[\|\mathtt{TRIP}\ x\|_{\mathtt{FREE}}]^{+}$.

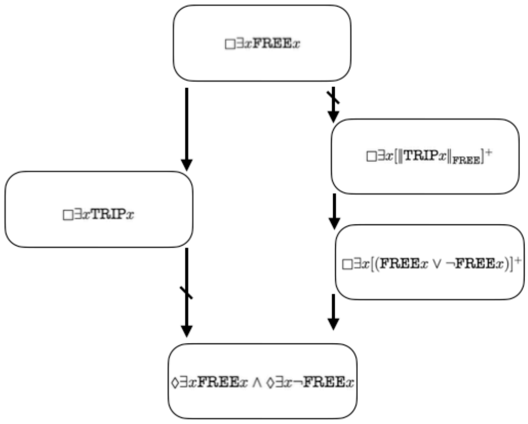


Figure 4.3 Solution to Asher's puzzle

The reinterpretation of $\mathtt{TRIP}$ can be processed as follows:

- Reinterpretation of $\mathtt{TRIP}$
  $\Box\exists x[\|\mathtt{TRIP}\ x\|_{\mathtt{FREE}}]^{+}$
  $\vDash \Box\exists x[(\mathtt{FREE}\ x \vee \neg\mathtt{FREE}\ x)]^{+}$
  $\vDash \Diamond\exists x\neg\mathtt{FREE}\ x$

The the details of the solution to Heim's example, which shares similarities with the solution to Asher's puzzle, are omitted. In Heim's example, the predicate "teach" is reinterpreted as "teach on Tuesday or teach on another days". Subsequently, the $\Box$-FC principle is applied to the reinterpreted predicate as follows:

- Reinterpretation of $\mathtt{TEACH}$
  $\Box\exists x[\|\mathtt{TEACH}\ x\|_{\mathtt{TTUE}}]^{+}$
  $\vDash \Box\exists x[(\mathtt{TTUE}_{\mathtt{TEACH}}\ x \vee \neg\mathtt{TTUE}_{\mathtt{TEACH}}\ x)]^{+}$
  $\vDash \Diamond(\neg\mathtt{TTUE}_{\mathtt{TEACH}}\ x)$

Also, if we think of the inferential pattern of the Heim's example to be (17), then it can be analyzed by $\Diamond$-$\mathtt{FC}$.

Table 4.3 presents a comparison between our proposal and Heim's and von Fintel's analyses of the puzzles.

Thus, in order to provide a consistent and coherent explanation of the puzzles, we

	Ross' Paradox	**Asher's Puzzle**	**Heim's example**
Von Fintel	Semantically valid	Pragmatically invalid	Presupposition failure
Heim	Semantically valid	Semantically invalid	Presupposition failure
WEM	Pragmatically invalid	Pragmatically invalid	Pragmatically invalid

Table 4.3 Comparing Predictions of Analyses

propose a uniform approach that treats them as instances of pragmatic failure. This proposal aims to address the deficiencies and inadequacies of previous accounts, and offers a more comprehensive and nuanced understanding of the role that pragmatic factors play in linguistic interpretation. Our approach is grounded in an analysis of the $\Box$- and $\Diamond$-free choice inferences that are commonly involved in generating these puzzles, and seeks to provide a more complete and accurate picture of how these inferences are processed and used in discourse, namely the reinterpretation of predicates. This proposal has the potential to contribute to a more general account of deontic paradoxes related to monotonicity and other linguistic phenomena. In the subsequent section, we explore the potential applications of this proposal and consider its broader implications for the study of language and communication.

4.5 Exploring Applications in Deontic Paradoxes

Deontic paradoxes related to upward monotonicity have long been a thorny issue in deontic logic and deontological ethics, posing difficult challenges to traditional frameworks based on principles such as the satisfaction of obligations and the avoidance of prohibitions. In attempting to resolve these paradoxes, philosophers and logician have proposed a variety of different approaches, ranging from relativistic and context-sensitive theories of obligation to rule-utilitarian frameworks that prioritize the greater good over individual duties, e.g. Geach (1956); Sorensen (1988); Horty (2007); Van Benthem and Liu (2018); Cariani (2013); Fox (2010); Von Fintel and Gillies (2010). The focus of these debates is whether the semantics of deontic operators are monotonic. We refrain from engaging in a semantic analysis of these paradoxes and instead focus on whether our proposal can be effectively applied to them. Our objective is to determine whether these paradoxes pose a genuine challenge to the upward monotonicity of deontic operators.

We will begin by presenting a list of paradoxes, followed by a detailed explanation of each.

(28) **Ross' Paradox** (Ross (1944))

a. Post this letter!

b. Post this letter or burn it!

(29) **Professor Procrastinate**: Professor Procrastinate is invited to review a book. He can do the review, but if he says yes, he would keep on postponing the task. So, although he can say yes and write the review, if he says yes, he would not write the review, which is the worst thing that can happen. (Jackson and Pargetter (1986))

a. Procrastinate ought to accept and write.

b. Procrastinate ought to accept.

(30) **Jump off** (Fox (2012))

a. You are obliged to jump off the bridge and land on the train.

b. You are obliged to jump off the bridge.

(31) **Gentle Murderer** (Forrester (1984))

a. It is obligatory that Jones not murder Smith.

b. If Jones murders Smith, it is obligatory that Jones murders Smith gently.

c. Jones murders Smith.

d. It is obligatory that Jones murders Smith gently.

e. It is obligatory that Jones murders Smith.

(32) **The Good Samaritan** (Prior (1958))

a. It is obligatory that Jones helps Smith who has been robbed.

b. It is obligatory that Smith be robbed.

It should be noted that the aforementioned paradoxes are not an exhaustive list of all the paradoxes related to monotonicity. There exists a wide range of such puzzles, and new ones are constantly being introduced. However, the examples presented above are representative of some typical types.

The deontic **Ross Paradox** can be analyzed similarly to its desiderative version (see Figure 4.2), wherein the application of free choice principle to (28-b) results in a conjunction of the permission to mail the letter and the permission to burn the letter. This paradox is related to deontic FC (Kamp (1974)).

The paradoxes labeled (29) and (30) appear to be analogous in that the conclusions in both cases do not seem to follow from their respective premises. Our analysis reveals that the premises express concrete obligations, such as "accept and write" and "jump off and land on the train," while the conclusions are more general, underspecific, simply "accept" and "jump off." These paradoxes are similar to Asher's puzzle (see Figure 4.3). As in the case of the predicate "free trip" (a predicate "trip" with the constraint of "free"), "write the review" and "land on the train" are constraints on the predicates "accept" and "jump off," respectively. The premises indicate that there are obligations to accept and jump off, but only if the constraints are satisfied. With our proposed reinterpretation, "accept" and "jump off" are understood to mean "accept and write or accept and don't write" and "jump off and land on the train or jump off and don't land on the train" respectively. By applying the free choice principle, we can derive the conclusions "Procrastinate is permitted to accept and not write" and "You are permitted to jump off the bridge and not land on the train." Thus, in both paradoxes the conclusions are infelicitous because they suffer from pragmatic failure.

In the case of (31)[1], the inference from (31-b) and (31-c) yields (31-d). However, the conclusion (31-e) derived from (31-d) contradicts the innocent assumption (31-a). This paradox bears similarities to Heim's example of conditional desire in (17), where the sentences "Jones ought to murder Smith gently" and "Jones ought to murder Smith" should be subject to the same condition, namely, "Jones is going to murder Smith". In this case, we can apply our analysis of reinterpretation: the predicate "murder Smith" would be reinterpreted as "murder Smith gently or murder Smith non-gently", and this reinterpretation should occur under the condition of Jones's intended murder. Therefore, the inferences can be rephrased as follows:

(33) If Jones murders Smith

a. Jones ought to murder Smith gently.
b. Jones ought to murder Smith.
c. Jones ought to murder Smith gently or murder Smith non-gently.
d. Jones is permitted to murder Smith non-gently.

By disjunctive reintepretation we interpret (33-b) as (33-c) and then by FC principle, we can derive the statement in (33-d), which is unacceptable. In this way the paradoxical nature of the conclusion (33-b) is explained as a case of pragmatic failure, since it allows

1 It is worth noting that a desiderative version of the Gentle Murderer paradox has been presented in Crnič (2011), which can be resolved as a presupposition failure using Heim's or von Fintel's dynamic analysis.

a inference (33-d) that is not justified by (33-a). It is crucial to note that it needs the inferences made under the same condition. This proposal is similar to Kratzer's theory in Kratzer (1991) which argued that (31-d) and (31-e) have the same modal base and source order, which is different from that of (31-a). In her analysis, when moving from (31-a) to (31-b), the modal base and ordering source have changed. In the modal base on which the assumption is predicated, all best worlds are non-murder-worlds. In contrast, in the modal base on which (31-b) is based, the best worlds are murder-worlds[1].

In summary, our analysis reveals that (i) the conclusion "Jones ought to murder Smith" is not in contradiction with the assumption "Jones ought not to murder Smith" due to their distinct underlying conditions, and (ii) the conditional "If Jones murders Smith, Jones ought to murder Smith" is intuitively unacceptable, as it can imply an objectionable permission of non-gentle murder. However, it also has a defensible aspect as it may convey a permissible meaning of gentle murder. Our theory is not intended to provide a comprehensive solution to this paradox; rather, we aim to demonstrate that our theory can be used to analyze aspects of this paradox in a meaningful way.

The Good Samaritan paradox has generated significant scholarly discussion and has been the subject of various proposals, such as those put forward by Aqvist (1967); Fraassen (1972); Kratzer (1991); Castañeda (1981). A comprehensive overview of these proposals can be found in Fox (2010). However, this paradox differs from others in our analysis. It has been argued that the following sentence is true:

(34) Jones helps Smith who has been robbed if and only if Jones helps Smith and Smith has been robbed.

Therefore, when the premise is paraphrased into standard deontic logic (SDL), it can be formalized as an obligation of a conjunction:

(35) $\texttt{[O]}(\mathrm{H}a \wedge \mathrm{R}a)$.

The symbol $\texttt{[O]}$ is used to represent the operator for obligation, while a is a constant representing Smith. The predicates $\texttt{H}$ and $\texttt{R}$ correspond to the actions of helping and robbing, respectively. Applying the principle of conjunction elimination to the premise in (35), we derive the formula in (35), which represents the obligation to rob Smith:

1 For more details of Kratzer's theory, refer to Kratzer (1991); Willer (2014).

(36) $\texttt{[O]}(\texttt{R}a)$

This conclusion is deemed unacceptable as there cannot be an obligation to rob. The paradox appears to resemble Asher's puzzle and Heim's example, where the elimination of conjunction (e.g. $\texttt{[Want]}(\texttt{FREE}\wedge\texttt{TRIP}) \Rightarrow \texttt{[Want]}\texttt{TRIP})$) results in a paradoxical conclusion, thus making the reinterpretation analysis seem applicable. Consider the pragmatically enriched version of (37).

(37) a. $\texttt{[O]}[((\texttt{R}a \wedge \texttt{H}a) \vee (\texttt{R}a \wedge \neg\texttt{H}a))]^{+}$

b. It is obligatory that Jones helps Smith who has been robbed **or** Jones doesn't help Smith who has been robbed.

c. $\texttt{[P]}(\texttt{R}a \wedge \texttt{H}a) \wedge \texttt{[P]}(\texttt{R}a \wedge \neg\texttt{H}a)$

d. It is permitted that Jones helps Smith who has been robbed, **and** it is permitted that Jones doesn't help Smith who has been robbed.

Formula (37-c) is derived from (37-a) by applying the $\Box$-FC principle, where the operator $\texttt{[P]}$ denotes permission. As shown by the inferences, not every conjunct in (37-d) seems unacceptable. The conjunct $\texttt{[P]}(\texttt{R}a \wedge \neg\texttt{H}a)$ conveys the permission for the act of not helping, which contradicts with the obligation of helping robbed people. However, $\texttt{[P]}(\texttt{R}a \wedge \texttt{H}a)$ expresses the permission of helping. So we at least identify a "positive" meaning that would support the affirmation of the conclusion (32-b).

The case under consideration is similar to that of the Gentle Murder paradox, which, even though its conclusion is counterintuitive, at least yields a coherent interpretation of "permission to murder gently" that is utterly acceptable. We can also formalize the paradox into a conditional version in a similar way to the case of the Gentle Murder, and then apply our pragmatic solution. For example, consider the following conditional version (also see Kratzer (1991)):

(38) If Smith has been robbed,

a. It is obligatory that Jones help Smith who has been robbed.

b. It is obligatory that Smith be robbed.

Applying the reinterpretation and free choice principle, we can derive the permission of not helping Smith. Despite the assumption that Smith has been robbed, it cannot accept the additional inferences from the conclusion "it is permitted that Jones doesn't help Smith who has been robbed".

A possible issue that may arise is the translation problem. When translating the formula $\mathtt{R}a \wedge \mathtt{H}a$ as "Smith has been robbed and Jones helps him", the conclusion "It is permitted that Smith has been robbed and Jones helps him" may appear infelicitous, as it seems to suggest that robbery is permissible. To address this issue, we may consider questioning the condition in (34). Another alternative is to adopt a different translation strategy: the relative clause "who has been robbed" should not be formalized within the scope of obligation. This is similar to the analysis for (19).

This example seems to have a unique feature, such as the special lexical items used. The actions described by the words robbery and murder are inherently in conflict with moral principles, regardless of any other constraints. If we replace these verbs with words that do not conflict with moral principles, the reasoning becomes less problematic. Consider the following example proposed by (Fox (2010, 2012)):

(39) **The Hygienic Cook** (Fox (2010))

a. It is obligatory that using a knife that has been cleaned.

b. It is obligatory that a knife has been cleaned.

The line of reasoning in (39), which shares the same structure as the Good Samaritan paradox, appears to be felicitous since the conclusion seems acceptable. We acknowledge that the issue of how to handle lexical items and their role in creating paradoxes requires further discussion, but this is outside the scope of the present chapter. We refrain from further exploring this matter and instead defer it to the reader for their own analysis and consideration.

4.6 Summary

Formulating a precise definition for the WEM presents a considerable challenge, given it refers to a class of empirical facts in natural language and there exists distinct accounts for these phenomena. Within the context of our theory, the most clear characterization of the effect may describe phenomena in which semantic inferences yield pragmatic consequences. Consequently, it is plausible to suggest that not only monotonicity reasoning, but also other forms of semantic inferences, have the potential to give rise to pragmatic effects.

In this chapter, we have discussed three puzzles that arise for monotonic semantics for desire verbs, and have proposed a uniform pragmatic account to address these puzzles. Another contributions of our proposal is that it shows how FC effects can be triggered by statements without overt disjunctions, since predicates in the sentences can be reinterpreted as disjunctive. This leads to additional inferences being drawn from monotonic reasoning, some of which are not semantically valid. To capture this reasoning, we have employed the logic QBSML extended with a reinterpretation function. Moreover, our account also sheds light on a number of classical deanotic paradoxes. A full analyses of the latter however must be left to another occasion. .

CHAPTER 5 A LOGIC FOR DESIRE BASED ON CAUSAL INFERENCE

In Chapter 4, our proposal of WEM theory provides evidence for the idea that these puzzles don't provide counter-examples against interpreting bouletic modals monotonically. In this chapter, we come up with a semantics for *want* and a logic corresponding the semantics. In our theoretical framework, the related monotonicity puzzles do not function as the determining factor in evaluating the monotonicity of the semantics for *want*. Consequently, our focus will be shifted towards examining the semantic feature of *want*, which determines the monotonicity of *want*.

5.1 Introduction

In many fields, ranging from logic to computer science, decision theory and semantics, the need for models to interpret *desire* has been recognized. This is not only because of the unique theoretical status of the concept of desire, but also because at the practical level, the agent's desires often provide a way for her to interact with the world. Accordingly, desire is usually interpreted from a perspective of decision theory, namely it is commonly modeled as a choice among options consistent with preference.

This preference-based approach has inspired many studies on desire. In Lang (1996); Lang et al. (2003), the authors propose formal systems of desire taking advantage of utility function on possible worlds. In the field of preference logic, a desire behavior is commonly regarded as an outcome of the entanglement of preferences and beliefs. Liu (2011a, 2007) give various logical approaches to combine preference with belief or other informational attitudes, and study both their static and dynamic structure. Some linguistic phenomena of desire also have been studied in a way ranking the modal base by adding a partial order function, such like Geurts (1998); Heim (1992); Villalta (2000). As introduced in Chapter 4, Stalnaker (Stalnaker (1984)) noted that "wanting something is preferring it to certain relevant alternatives, the relevant alternatives being those possibilities that the agent believes will be realized if he does not get what he wants." Heim, in Heim (1992), reformulates Stalnaker's account of want as follows: ϕ is desired if and only if the closest ϕ-alternatives (with respect to the actual world) are more preferred to the closest $\neg\phi$-

alternatives according to the agent. In view of this analysis, in addition to the preference structure, the conditional also plays a role in interpreting desire.

Following the preference-based approach, and being inspired by Stalnaker's account, we attempt to model desire based on causal inferences in this chapter. The intuition is that a desire for an object is determined by the effects it causes, and causality will explicitly play an important role in accounting for the casual relation. To do so, we will focus on the method of combining causality and preference structure in the formal framework. To clarify our motivation and explain why we need to extend the model from a causal perspective, let us take an example first:

Example 1: *Robin has been in charge of an important project in the company recently, and the tremendous pressure made him sleepless every night. Robin is a health-conscious person, and poor sleep will make him sluggish the next day which bothers Robin most. So having a good sleep is what he needs most at the moment. Therefore, on this sleepless night, Robin wants to take some sleep pills to help him sleep well, even if these pills have some side effects*

For the simplicity of the discussion, let us abbreviate the proposition "pills are taken" as P, and "Robin sleeps well" as S. The four possible situations in this example can be named as below: $P\neg S$ be the possible world in which Robin takes the pills but still be sleepless; $\neg PS$ is the possible world in which Robin takes no pills and sleeps well; PS is the possible world in which Robin takes sleeping pills and sleeps well; $\neg P\neg S$ is the possible world in which Robin is sleepless without taking any pills.

In Robin's example, the preference structure can be characterized by the following ordering: $P\neg S \prec \neg P\neg S \prec PS \prec \neg PS$ where $\prec$ means "not as good as".

It is intuitive to say that Robin desires to take sleeping pills, although (at least from the perspective of many classical theories of preference) he does not prefer to do so. According to the theory of "preference lifting" (e.g. Halpern (2017); Liu (2011a); Shi and Sun (2021)), "taking sleeping pills" is not preferred under the preference structure above, because there is a possible world $P\neg S$ in which Robin takes the sleeping pills while it is worse than any possible world in which Robin does not take the pills.

If we follow Stalnaker's analysis of desire, a possible explanation for Robin's desire of P without his preferring P will be that: in the evaluation of Robin's desire for P, the possible world $P\neg S$ should not be taken into the consideration, as it should not be seen as "relevant alternative" in this context. However, it is insufficient to say that $P\neg S$ is an irrelevant alternative because it is abnormal/with low possibility. To illustrate this let us consider the following variation of Robin's example:

Example 2: *Robin's company offers him a chance to study in the Netherlands and he attaches great importance to it. But his savings are small, so he doesn't want to pay for the tuition. Unfortunately, due to the epidemic, he cannot go abroad. Even after paying the tuition, he can only stay at home and take online lessons.*[1]

Let us abbreviate "paying the tuition" as P' and " study in the Netherlands" as S'. The preference ordering is $P'\neg S' \prec \neg P'\neg S' \prec P'S' \prec \neg P'S'$. The abnormal worlds are $P'\neg S'$ and $\neg P'S'$. The preference structure of this variant is isomorphic to the original one. If we apply the previous argument which excludes the abnormal worlds, and only compare $\neg P'\neg S'$ with $P'S'$, then we have to conclude that Robin desires to pay tuition, which is a counter-intuitive conclusion.

Therefore the question is why $P\neg S$ are irrelevant in the first example while $P'\neg S'$ are relevant in second one. However preference structures do not suffice to solve it. As we have seen, both examples have the same structure of preference, whereas the agent has different desires.

According to our diagnosis, the causal feature differentiate the two examples: in the first example, taking pills (P) causes good sleep (S), while in the second example paying tuition (P') does not cause studying Netherlands (S').

To tackle these questions, the causal information need to be taken into account the preference structure. In general, our account of desire still follows the framework proposed by Stalnaker. Our model of desire works as follows: the causal information in the model explicitly predicts which possible worlds are the "relevant alternatives"; And the preference structure tells whether the "relevant alternatives" are preferred by the agent

This chapter is organized as follows. In Sec.5.2, we will first review the classical methods of modelling preference and causality. Then, in Sec.5.3, by combing the idea of these two modelling methods, we propose a new model to interpret desire based on the preference and causality. A logic system will be given in Sec.5.4. In Sec.5.5 we will show that the our logic system enables us to discuss some properties such like the entanglement between desire and causality.

1 The original version of this example comes from Maria Aloni (pc).

5.2 Modelling methods for preference and causality

As mentioned above, we propose an interpretation of desire by combining causality and preference structures. In this section, as a preliminary, we look at the approaches to modelling preferences and causality respectively. For preferences, we mainly focus on the static structure concerning *betterness* relations and *priority* ordering (see Van Benthem et al. (2006); de Jongh and Liu (2009)). And for causality, we will introduce the interventionist approach to causal model (see Pearl (1995, 2000); Halpern (2000)). In the section 5.3, we will combine these two modelling approaches to construct a new model for desire.

5.2.1 Modelling methods for preference

Preference plays an important role in decision making. In this chapter our main focus is on the preference over states of affairs, which indicates the preference arises from comparisons between outcomes, actions, or situations, etc. States of affairs are commonly represented as propositions in logical systems.

5.2.1.1 Betterness model

From the literature following Von Wright's tradition, such as Boutilier (1994); Halpern (1997), the preference between propositions can be represented as a binary relation over possible worlds in Kripke's semantics. This relation, in Van Benthem and Liu (2007); Van Benthem et al. (2006); Liu (2007), is formalized by a *betterness order* of possible worlds in a *betterness model*.

Definition 33 (Betterness model)**:** *A betterness model is a tuple $M = (W, \preceq, V)$ where W is a set of possible worlds, $\preceq$ is a reflexive and transitive relation (the 'betterness' pre-order) over these worlds, and V is a valuation assigning truth values to proposition letters at worlds.*

For every possible worlds w_1 and w_2, $w_1 \preceq w_2$ can be read as 'w_2 is at least as good as w_1'. In addition, if $w_1 \preceq w_2$ but not $w_2 \preceq w_1$, then w_2 is strictly better than w_1, written as $w_1 \prec w_2$.

The intuition of modeling preference over propositions using betterness relation is that a proposition ϕ is preferred in a world w whenever ϕ is true in at least one world which is considered at least as good as w. More discussion about preference on betterness order can be found in Liu (2007, 2011b). In what follows, we will introduce a method of deriving the betterness relation among possible worlds from priority ordering, which explain how the preference comes into being.

5.2.1.2 Priority of propositions

In a betterness model, preference is characterized as a comparison between two alternatives according to the betterness order. So, to discuss the origin of preference over propositions, it needs to consider how to draw the betterness order, assuming there is a basic source where it can be derived. There are various ways to model the process of producing preference (see Coste-Marquis et al. (2004) to have an overview of the discussion). In this chapter, we adopt the method proposed in de Jongh and Liu (2009), where the betterness order of worlds is derived from a basic source in which the propositions are ordered according to their importance. The insight comes from *Optimality Theory* in linguistics (see Prince and Smolensky (2008); McCarthy (2007)), which is a model of explaining how grammars are structured.

In optimality theory, it provides an approach of choosing an optimal solution among alternatives by applying a set of conditions (so-called constraints) that are ranked for importance. Inspired by this process, in de Jongh and Liu (2009), the authors propose that preference can be drawn from a *priority order* consisting of propositions, which is linked to graded semantics and considered to be linearly ordered. In the following, we present the definitions that come directly from de Jongh and Liu (2009); Van Benthem et al. (2014).

Definition 34: *[Priority order] A priority order $\mathcal{G} = \langle \Phi, \ll \rangle$ is a strictly partially ordered set of propositions in basic betterness language.*

A priority order where $\ll$ is a strict linear order is called a priority sequence, which is represented as a sequence $\langle \phi_1, ..., \phi_n \rangle$ of propositional formulas. Intuitively, the affairs represented by ϕ_n are strictly ranked based on their importance. The following definition shows how a given priority order determines a betterness ordering of possible worlds:

Definition 35 (Betterness from a priority order)**:** *Let $\mathcal{G} = \langle \Phi, \ll \rangle$ be a priority order, W a non-empty set of possible worlds. The derived betterness relation $\preceq_{G} \subseteq W^2$ is defined as follows*[1]*:*

$$w \preceq_{\mathcal{G}} w' := \forall \phi \in \mathcal{G}\colon w \in [\![\phi]\!] \Rightarrow w' \in [\![\phi]\!]$$

Assume $[\![\cdot]\!]$ is a valuation function: $\Phi \to 2^W$. Accordingly, the possible worlds in W are ordered according to which elements of the priority order they satisfy. If a possible

1 There is a syntactically different, but semantically equivalent format of priority order, defined in Liu (2011b).

world w satisfies a proposition in the Φ, then it also satisfies all propositions that are no more important than it. And we suppose that a possible world w satisfies a proposition ϕ if and only if ϕ is true on w. Based on Definition 35, the following properties of the derived betterness relation $\preceq_{\mathcal{G}}$ hold:

Fact 14 (Properties of $\preceq_{\mathcal{G}}$)**:** *Let $\mathcal{G} = \langle \Phi, \ll \rangle$ be a priority order, then*

- *The relation $\preceq_{\mathcal{G}}$ is a pre-order whose strict part $\prec_{\mathcal{G}}$ is upward well-founded;*
- *For arbitrary i, j if $\phi_i \ll \phi_j$, then for all worlds $w \in [\![\phi_i]\!]$, $w' \in [\![\phi_j]\!]$: $w \preceq_{\mathcal{G}} w'$;*
- *For arbitrary i, j if $\phi_i \ll \phi_j$, then for all worlds $w \in [\![\phi_i \wedge \neg\phi_j]\!]$, $w' \in [\![\phi_j]\!]$: $w \prec_{\mathcal{G}} w'$*

In this chapter, we will use similar strategies to derive preference over propositions. Desire, in our proposal, is drawn by the preference relation combined with causal reasoning. In the next section, we will introduce the formal approach of modelling causality.

5.2.2 Modelling methods for causality

In recent years, there is a development of formal approaches to causal reasoning and causal learning, based on the work of structural equation causal models in Pearl (1995). The logic based on causal models, which goes back to Galles and Pearl (1998) and was further developed in Halpern (2000); Pearl (2000); Briggs (2012). We will briefly introduce the causal modelling approach on which our study of desire will be built.

Causal models consists of two parts: a finite set of causal variables, and a set of structural equations which represents the causal relationships among these variables. A (structural equation) causal model can be formally defined as follows[1]:

Definition 36 (Causal model)**:** *A causal model is a tuple $\langle \mathcal{V}, \mathcal{F} \rangle$. $\mathcal{V}$ is the set of causal variables. Let Σ be all possible values of $\mathcal{V}$, $\mathcal{F}$ is a collection of functions $\{f_X\}_{X \in \mathcal{V}}$ with $f_X : (\mathcal{V} \to \Sigma) \to \Sigma$.*

Σ represents the range of causal variables. We call a (partial) function from $\mathcal{V}$ to Σ as a (partial) assignment to $\mathcal{V}$. In some systems, such as the logic system in Ibeling and Icard (2020), Σ is assumed to be $\{0, 1\}$, so that the variables can be interpreted as atomic propositions. The intuition behind $\mathcal{F}$ (which is called the set of structural functions) is

1 Notice that there are various way to formalize causal models, and here we just choose one way among them. In Pearl, Galles and Halpern's original work, the causal variables are sorted into exogenous variables and endogenous variables, depending on whether get "influenced" by other variables. In this chapter we will use a more simplified version of causal models, as in Ibeling and Icard (2020), in which there is no such distinction.

that: $\mathcal{F}$ is a set of causal rules such that for each variable V, $f_V \in \mathcal{F}$ gives the information about how other variables determine the value of V. So $\mathcal{F}$ intuitively represents the causal rules in the actual world. It could be the case that V is influenced by no other variables, and in this case f_V is a constant function standing for V's actual value.

It should be noticed that, for each structural function $f_X \in \mathcal{F}$, though it is a function that takes a full assignment $\mathcal{A}$ of $\mathcal{V}$ as the input, it could be the case that some variables do not play any role in determining the value of $f_X(\mathcal{A})$. For instance, it is possible that $f_X(\mathcal{A}_1) = f_X(\mathcal{A}_2)$ always holds whenever $\mathcal{A}_1$ and $\mathcal{A}_2$ agree on the value of every variable in $\mathcal{V}\backslash\{Y\}$. In this case, we say Y is *not* a causal parent of X. Given a causal model $\langle\mathcal{V},\mathcal{F}\rangle$, we can generate a graph whose nodes represent causal variables and edges represent the relation of causal parent (X is a successor of Y in the graph if and only if Y is a causal parent of X). Such a graph is known as a causal graph. We usually assume a causal graph should be acyclic, which means there is no causal loops. A causal model is called a recursive causal model if and only if its causal graph is acyclic. In this chapter, all the causal models are assumed to be recursive.

If $\mathcal{F}$ is recursive, then there is exactly one assignment $\mathcal{A}$ to $\mathcal{V}$ that complies with all causal rules $\{f_X\}_{X\in\mathcal{V}}$, which means: for any $X \in \mathcal{V}$, $f_X(\mathcal{A}) = \mathcal{A}(X)$[1]. The unique assignment to $\mathcal{V}$ that complies with $\mathcal{F}$ is called *The unique solution to* $\mathcal{F}$ (write $\mathcal{A}^{\mathcal{F}}$ when $\mathcal{V}$ is clear in the context). Given a causal model $\langle\mathcal{V},\mathcal{F}\rangle$, $\mathcal{A}^{\mathcal{F}}$ intuitively represents the actual state of the variables.

The language describing the causal models, in the most well-known version, is given by extending the propositional language (whose atomic sentences are $X = x$ saying the value of X is x) with counterfactuals $\vec{X} = \vec{x} \mathrel{\Box\!\!\rightarrow} \phi$ which should be read as a counterfactual conditional: "if the variables in $\vec{X}$ were set to the values $\vec{x}$, respectively, then ϕ would be the case". See below:[2]

Definition 37 (Basic causal language $\mathcal{L}_C$)**:** *Formulas ϕ of the language $\mathcal{L}_C$ for variable $\mathcal{V}$ are given by*

$$\varphi ::= Y{=}y \mid \neg\phi \mid \phi\wedge\phi \mid \vec{X}{=}\vec{x} \mathrel{\Box\!\!\rightarrow} \phi$$

for $Y \in \mathcal{V}$, $y \in \Sigma$ (Σ is the range of $\mathcal{V}$) and $\vec{X}{=}\vec{x}$ a (possibly partial) assignment on $\mathcal{V}$.

1 This property has been shown in Halpern (2000).

2 It should be noticed that there are many differences in the languages for causal models in various formal systems. In some literature such as Halpern (2013), a counterfactual $\vec{X} = \vec{x} \mathrel{\Box\!\!\rightarrow} \phi$ will be written as $[\vec{X} = \vec{x}]\phi$. In Ibeling and Icard (2020), $X{=}1$ is written as X and $X{=}0$ is written as $\neg X$. Some causal language has a restriction that only variables influenced by other variables (endogenous variables) can appear in the antecedent, while this restriction is dropped in some other systems, such as Barbero et al. (2020) and Ibeling and Icard (2020).

The truth condition of a counterfactual $(\vec{X} = \vec{x}) \,\Box\!\!\rightarrow \phi$ is that the consequent ϕ holds at the causal model results from an "intervention":

Definition 38 (Semantics of counterfactuals based on intervention)**:** *Let $\langle \mathcal{V}, \mathcal{F} \rangle$ be a causal model with $\mathcal{F} = \{f_X\}_{X \in \mathcal{V}}$.*

- *$\langle \mathcal{V}, \mathcal{F} \rangle \vDash Y = y$ iff $\mathcal{A}^{\mathcal{F}}(Y) = y$ where $\mathcal{A}^{\mathcal{F}}$ is the unique solution to $\mathcal{F}$.*
- *The Boolean cases are defined as usual.*
- *$\langle \mathcal{V}, \mathcal{F} \rangle \vDash (X_1 = x_1, ..., X_n = x_n) \,\Box\!\!\rightarrow \phi$ iff $\langle \mathcal{V}, \mathcal{F}_{X_1 = x_1, ..., X_n = x_n} \rangle \vDash \phi$ where $\mathcal{F}_{X_1 = x_1, ..., X_n = x_n}$ is the set of functions results from replacing $f_{X_1}, ..., f_{X_n}$ by constant functions whose outputs are $x_1, ..., x_n$ respectively.*

The intuition behind the truth condition of counterfactual $(X_1 = x_1, ..., X_n = x_n) \,\Box\!\!\rightarrow \phi$[1] is that: we are first forcing the value of $X_1, ..., X_n$ to be $x_1, ..., x_n$ (replacing their structural functions by constant functions whose output is $x_1, ..., x_n$), and preserve the causal dependence that is not involved in the antecedent as much as possible, then check whether the consequent ϕ holds in this intervened model.

Halpern (2000) provides a logic system based on this interventionist semantics of counterfactuals.

5.3 Modelling desire by combining the two approaches

In this section, we will build a model that combining the two approaches introduced in Sec.5.2. In addition to representing the preference of the agent, the new model characterizes the causal dependencies.

The priority order, which comes from the modelling method for preference, plays a role in our account. In the example of sleeping pills, we need to explain Robin's choice between a world in which no pills are taken but sleepless, and a world in which Robin sleeps well but with pills taken, as well as the reason behind this choice: it is more important to sleep well than to take pills. In order to model the different importance of propositions to the agent, the priority order defined in 34 can be drawn on. The difference between our approach and the proposals in de Jongh and Liu (2009); Van Benthem et al. (2014) is that our priority order is based on causal variables instead of propositions. Also, we assume that the priority relation is total, therefore $\langle \mathcal{V}, \ll \rangle$ is defined to be a priority sequence.

1 The truth condition for Boolean conjunctions of atomic sentences can be seen as a special case of the truth condition for counterfactuals $(X_1 = x_1, ..., X_n = x_n) \,\Box\!\!\rightarrow \phi$ in which $X_1 = x_1, ..., X_n = x_n$ is an empty sequence.

$V_i \ll V_k$ means V_k is more important than V_i. In this way, we can represent "sleeping well is more important than other factors for Robin" by $P \ll S$.

In addition, combining the perspectives of causality and preference enables us to distinguish two kinds of desire: *intrinsic desire* and *extrinsic desire*. People desire something intrinsically will desire it for its own sake. While extrinsic desire, which is also called instrumental desire, is a desire depending on the effects it causes. Consider the example of sleeping pills again. Robin has an intrinsic desire for a good sleep, but not for taking sleeping pills. A good sleep is desirable no matter whether the good effect of it is taken into consideration. But the desire of taking sleeping pills is not intrinsic: if sleeping pills do not cause any desirable things, such as better sleep, then nobody would like to take it. Such a distinction is important for our analysis of desire, therefore we will extend the model by adding a intrinsic preference $In : \mathcal{V} \to \{0, 1\}$ representing the intrinsic desire of the agent. For each causal variable X, $In(X) = 1$ means the agent has the intrinsic desire of $X = 1$ and $In(X) = 0$ means the opposite case. In Robin's case, $In(S) = 1$ and $In(P) = 0$.

Following this idea, we build a desire-causality model by extending a causal model with a priority relation $\ll$ and a intrinsic preference function In. A desire-causality model is formally defined as:

Definition 39 (Desire-Causality Model)**:** *A desire-causality model based is a tuple $\mathcal{M} = \langle \mathcal{V}, \mathcal{F}, \ll, In \rangle$.*

- *$\mathcal{V}$ is the set of causal variables.*
- *$\mathcal{F}$ is a collection of functions $\{f_X\}_{X \in \mathcal{V}}$ with $f_X : (\mathcal{V} \to \{0, 1\}) \to \{0, 1\}$, and $\mathcal{F}$ is assumed to be recursive.*
- *$\ll$ is a strict total order over $\mathcal{V}$.*
- *In is a function from $\mathcal{V}$ to $\{0, 1\}$*

If $In(X) = x$, then we say $X = x$ is a preferred proposition. Let $X = x$ and $Y = y$ be two preferred propositions, we say $X = x$ is prior to $Y = y$ if $Y \ll X$. Each full assignment to $\mathcal{V}$ can be seen as a possible world, as it is a complete description of a situation in terms of the value of all variables. Let $\mathcal{A}$ be a full assignment, if $\mathcal{A}(X) = In(X) = x$, then we say $\mathcal{A}$ satisfies the intrinsic preference $X = x$.

Based on the priority sequence and the intrinsic preference function, we can derive the betterness relation between possible worlds (full assignments). The evaluation criteria of betterness are: $\mathcal{A}_1$ is better than $\mathcal{A}_2$ whenever there is a primarily preferred proposition

$X = x$ satisfied by $\mathcal{A}_1$ but not by $\mathcal{A}_2$, and $X = x$ is prior to every preferred proposition satisfied by $\mathcal{A}_2$ but not by $\mathcal{A}_1$.

Definition 40 (Betterness derived from priority sequence)**:** *Let $\mathcal{M} = \langle \mathcal{V}, \mathcal{F}, \ll, In \rangle$ be a desire-causality model, the betterness relation derived from $\mathcal{M}$, write $\preceq_{\mathcal{M}}$, is a binary relation over $2^{\mathcal{V}}$ (the set of all possible assignment to $\mathcal{V}$) defined as below:*

For any $\mathcal{A}_1, \mathcal{A}_2 \in 2^{\mathcal{V}}$, $\mathcal{A}_2 \preceq \mathcal{A}_1$ whenever: $\mathcal{A}_1 = \mathcal{A}_2$ or there is X such that $\mathcal{A}_1(X) = In(X) \neq \mathcal{A}_2(X)$ and for every Y with $\mathcal{A}_1(Y) \neq \mathcal{A}_2(Y) = In(Y)$, $Y \ll X$

$\prec_{\mathcal{M}}$ refers to the strict part of $\preceq_{\mathcal{M}}$: $\mathcal{A}_1 \prec_{\mathcal{M}} \mathcal{A}_2$ whenever $\mathcal{A}_1 \preceq_{\mathcal{M}} \mathcal{A}_2$ and $\mathcal{A}_2 \npreceq_{\mathcal{M}} \mathcal{A}_1$

Let $\mathcal{M} = \langle \mathcal{V}, \mathcal{F}, \ll, In \rangle$ be a desire causality model, the betterness relation on $\mathcal{M}$ satisfies the following properties:

Fact 15 (Properties of betterness relation in desire-causality models)**:**

- *For every i, j if $X_i \ll X_j$, then for all full assignments $\mathcal{A}(X_i) = In(X_i)$, $\mathcal{A}'(X_j) = In(X_j)$: $\mathcal{A} \preceq_{\mathcal{M}} \mathcal{A}'$;*
- *For every i, j if $X_i \ll X_j$, then for all worlds $\mathcal{A}(X_i) = In(X_i)$ and $\mathcal{A}(X_j) \neq In(X_j)$, $\mathcal{A}'(X_j) = In(X_j)$: $\mathcal{A} \prec_{\mathcal{M}} \mathcal{A}'$*

We can find a clear correspondence between Fact 14 and Fact 15, if we think of a full assignment $\mathcal{A}$ as a possible world w, and $\mathcal{A}(X_i) = In(X_i)$ as the counterpart of $w \in [\![\phi_i]\!]$.

Definition 41 (Language $\mathcal{L}_{DC}$)**:** *Formulas φ of the language $\mathcal{L}_{DC}$ based on $\mathcal{S}$ are given by*

$$\phi ::= X{=}x \mid \neg\phi \mid \phi \wedge \phi \mid (\vec{X}{=}\vec{x}) \mathrel{\Box\!\!\rightarrow} \phi \mid \mathbb{D}(\vec{X}{=}\vec{x}) \mid X < Y$$

where the variables in $\vec{X}$ are distinct, namely they are different from each other.

$\mathcal{L}_{DC}$ is the language that extends $\mathcal{L}_C$ with a desire-operator: $\mathbb{D}(X_1{=}x_1, ..., X_n{=}x_n)$ represents the desire that wanting $\vec{X}$ to be X_1,...,X_n to be $x_1, ..., x_n$ respectively. The meaning of $(X_1{=}x_1, ..., X_n{=}x_n) \mathrel{\Box\!\!\rightarrow} \phi$ is the same as in the basic causal language: it denotes the counterfactual conditional which means if $X_1{=}x_1$,...,$X_n{=}x_n$ were true, then ϕ would be the case.[1] For simplicity, we will also write $X_1{=}x_1 \wedge ... \wedge X_n{=}x_n$ as $\vec{X}{=}\vec{x}$ when the context allows.

Then we need to consider the semantic of $\mathcal{L}_{DC}$. The semantics of the boolean connectives and counterfactuals should be the same as in $\mathcal{L}_C$. The semantics of $<$ is straightforward, that is: $X < Y$ holds at $\mathcal{M}$ iff Y is prior to X according to the prior sequence

1 When $\vec{X}$ is an empty sequence, we will not distinguish $\vec{X} = \vec{x} \mathrel{\Box\!\!\rightarrow} \phi$ and ϕ in our language.

in $\mathcal{M}$. For the meaning of desire, our treatment is similar to Stalnaker's thesis of want: $X = x$ desired if and only if the closest $X = x$-alternative with respect to the actual world are more preferred to the actual world (where $X \neq x$) according to the agent.

As we have mentioned before, a full assignment to all causal variables can be seen as a complete description of a possible situation. Therefore, to find the closest $X = x$-world is to find the closest assignment that assigns X with value x, with respect to the actual assignment.

The assignment results from an intervention $X = x$ at the actual assignment is a good candidate for the closest $X = x$-alternative. Intuitively, $\mathcal{A}^{\mathcal{F}_{X=x}}$ results from the minimal change on the actual assignment that counterfactually making $X = x$ true: (i) $\mathcal{A}^{\mathcal{F}_{X=x}}(X)$ must be x (ii) only the value of the variables that are directly or indirectly dependent on X will be updated according to the new value after the intervention (iii) the values of other variables can remain the same as before (because any causal influence on X has been removed by the intervention, they are still consistent with the new value of X). Therefore, $\mathcal{A}^{\mathcal{F}_{X=x}}$ can be seen as the "closest alternative" where $X = x$ is forced to be true.

Therefore, we interpret "desire X=x" as: the assignment results from the intervention counterfactually making X=x true are better than the actual assignment, according to the betterness relation derived from the model.

Definition 42 (Semantics of $\mathcal{L}_{DC}$)**:** *Let $\langle \mathcal{V}, \mathcal{F}, \ll, In \rangle$ be a desire-causality model, the truth condition of the formulas in $\mathcal{L}$ are given by (boolean cases are defined as usual):*

- $\langle \mathcal{V}, \mathcal{F}, \ll, In \rangle \models X{=}x$ *iff* $\mathcal{A}^{\mathcal{F}}(X) = x$ *where* $\mathcal{A}^{\mathcal{F}}$ *is the unique solution to* $\mathcal{F}$
- $\langle \mathcal{V}, \mathcal{F}, \ll, In \rangle \models (\vec{X}{=}\vec{x}) \,\Box\!\!\rightarrow \phi$ *iff* $\langle \mathcal{V}, \mathcal{F}_{\vec{X}=\vec{x}}, \ll, In \rangle \models \phi$
- $\langle \mathcal{V}, \mathcal{F}, \ll, In \rangle \models \mathbb{D}(\vec{X}{=}\vec{x})$ *iff* $\mathcal{A}^{\mathcal{F}} \prec_{\mathcal{M}} \mathcal{A}^{\mathcal{F}_{\vec{X}=\vec{x}}}$*, where* $\mathcal{A}^{\mathcal{F}}$ *is the unique solution to* $\mathcal{F}$ *and* $\mathcal{A}^{\mathcal{F}_{\vec{X}=\vec{x}}}$ *is the unique solution to* $\mathcal{F}_{\vec{X}=\vec{x}}$
- $\langle \mathcal{V}, \mathcal{F}, \ll, In \rangle \models X < Y$ *iff* $X \ll Y$

The truth conditions of the atomic sentences and counterfactuals are the same as in the basic causal language. The desire-causality model results from intervening on $\langle \mathcal{V}, \mathcal{F}, \ll, In \rangle$ is defined as $\langle \mathcal{V}, \mathcal{F}_{\vec{X}=\vec{x}}, \ll, In \rangle$: it is natural to assume that intervention only changes the causal rules and actual state of variables, but does no modification on the agent's priority sequence and preference over single variables.

The intuition behind the desire operator $\mathbb{D}$ is that, the actual state is less preferred than the state after the intervention forcing the value of $\vec{X}$ to be $\vec{x}$ ($\mathcal{A}^{\mathcal{F}}$ can be seen as the actual state, $\mathcal{A}^{\mathcal{F}_{\vec{X}=\vec{x}}}$ can be seen as the state after the intervention forcing the value of $\vec{X}$ to be $\vec{x}$).

Consider the example of *Robin's desire of taking sleeping pills* again. Let P stand for 'taking sleeping-help pills' and S stand for 'having a good sleep'. Recall that in this example, good sleep is primarily desired, but sleeping pills are not (we assume nobody wants to take pills if it does not bring any good effect), so $In(P) = 0$, $In(S) = 1$. In addition, sleeping well is ranked as more important in the priority sequence, which can be modeled by $P \ll S$. This model can be described by Figure 5.1, in which the betterness relation derived from the intrinsic desire and priority relation is represented by solid arrow.

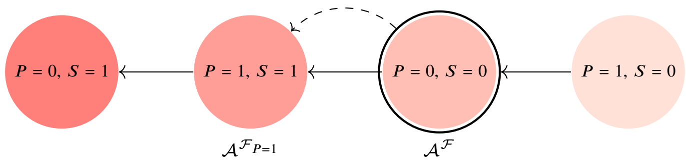


Figure 5.1 Robin's desire of taking sleeping pills

The circle with black edge represents the actual state; the solid arrow is the order $\preceq_{\mathcal{M}}$ derived from the intrinsic desire and priority relation and the dashed arrow is the counterfactual transition by setting the value of P to 1. [1]

For the structural functions in $\mathcal{F} = \{f_P, f_S\}$, f_P is a constant function whose output is 0 (because in the actual world, Robin has not taken sleeping pills at the moment) and f_S is a function such that $f_S(\mathcal{A}) = \mathcal{A}(P)$ (namely $S = 1$ whenever $P = 1$). As a result, it can be calculated based on $\mathcal{F}$ that the actual state $\mathcal{A}^{\mathcal{F}}$ is $(P{=}0, S{=}0)$, and the state results from intervening on the actual state $\mathcal{A}^{\mathcal{F}_{P=1}}$ is $(P{=}1, S{=}1)$. The dashed arrow in Figure 5.1 represents the counterfactual transition from the actual state to the state results from intervention making $P = 1$.

As can be observed in Figure 5.1, the actual state $\mathcal{A}^{\mathcal{F}}$ is less preferred than the state $\mathcal{A}^{\mathcal{F}_{X=1}}$ results from intervention, therefore the formula $\mathbb{D}(P = 1)$ is true in this desire-causality model according to Definition 42.

Now we apply the same way of modelling to Robin's desire of studying in the Netherlands. Let $N{=}1$ stands for studying in Netherlands and $N{=}0$ for not; And $T{=}1$ stands for paying tuition fee and $T{=}0$ for not. The corresponding desire-causality model is illustrated in Figure 5.2

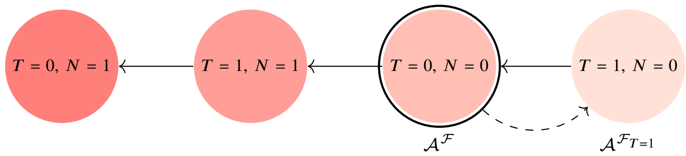


Figure 5.2 Robin's desire of tuition

The structural function for T, namely f_T, is a function such that $f_T(\mathcal{A}) = \mathcal{A}(N)$

(namely $T = 1$ whenever $N = 1$); while the structural function for N is a constant function whose output is 1 (because in the actual world Robin is not studying Netherlands). We can check that, according to Definition 42, the actual state $\mathcal{A}^{\mathcal{F}}$ is $(T{=}0, N{=}0)$ and the state results from forcing the value of T to be 1 is $\mathcal{A}^{\mathcal{F}_{T=1}} = (T{=}1, N{=}0)$. As shown in Figure 5.2, we can check that $\mathbb{D}(T{=}1)$ does not hold. The result fits our intuition that Robin does not desire paying for the tuition.

5.4 The logic for desire-causality models

As we mentioned in Section 2, Halpern (2000, 2013); Briggs (2012) provide a logic system L_C based on $\mathcal{L}_C$ with respect to (recursive) causal models. The axioms of L_C consists of all propositional tautologies, the MP rule, and A_1 to $\mathrm{A}_{[][]}$ in Table 5.1[1].

In this chapter, we extend the language $\mathcal{L}_C$ with the desire component. So we need an extended axiom system for the extended language $\mathcal{L}_{DC}$ with respect to desire-causality models. The new system, whose axiom and rules appear in Table 5.1, is given by extending the L_C with A_7 to A_{17} and generalization rule. We will name it as L_{DC}.[2].

The soundness of A_1 to A_6 has been proven in Halpern (2000) (In a slightly different form). The soundness of $\mathrm{A}_\neg$, $\mathrm{A}_\wedge$ and is straightforward, that is: intervention is deterministic, so Boolean operators can be distributed into the consequence. $\mathrm{A}_{[][]}$ means that, if two interventions are applied sequentially, the later intervention will over-wright the previous intervention.

The intuition behind A_7 is that if all the causal connection among variables is removed (A full intervention $\vec{V} = \vec{v}$ removes all causal connections in the model as each structural function is replaced by a constant function), the desire of Y is independent of the status of other variables. The intuition behind A_8 is that anything that already holds in the actual world will not be desired; A_9 indicates that if a situation A is desired conditional on X, then it is equivalent to say that the situation is also desired conditional on A together with X's consequence. A_{10} indicates that if X is desired, it is equivalent to say that X together

1 Because of the difference in notation, the original forms of the axioms in Halpern (2000, 2013); Briggs (2012) are rather different from what they look like in Table 5.1. For the readability, we reformulate their axioms.

2 In Table 5.1, $X \rightsquigarrow V$ is an abbreviation of the following formula:

$$\bigvee_{\substack{\vec{Z} \subseteq \mathcal{V} \setminus \{X, V\},\\ \{x_1, x_2\} \subseteq \{0,1\},\ x_1 \neq x_2,\\ \{v_1, v_2\} \subseteq \{0,1\},\ v_1 \neq v_2}} \vec{Z}{=}\vec{z}, X{=}x_1 \mathrel{\Box\!\!\rightarrow} (V{=}v_1) \wedge \vec{Z}{=}\vec{z}, X{=}x_2 \mathrel{\Box\!\!\rightarrow} (V{=}v_2).$$

P	φ	for φ an instance of a propositional tautology
MP	From φ_1 and $\varphi_1 \to \varphi_2$ infer φ_2	
Generalization	From ϕ infer $\vec{X} = \vec{x} \,\Box\!\!\rightarrow \phi$	
A_1	$\vec{X}{=}\vec{x} \,\Box\!\!\rightarrow Y{=}y \to \neg\vec{X}{=}\vec{x} \,\Box\!\!\rightarrow Y{=}y'$	for $y, y' \in \{0,1\}$ with $y \neq y'$
A_2	$\bigvee_{y\in\{0,1\}} \vec{X}{=}\vec{x} \,\Box\!\!\rightarrow Y{=}y$	
A_3	$\big(\vec{X}{=}\vec{x} \,\Box\!\!\rightarrow (Y{=}y) \wedge \vec{X}{=}\vec{x} \,\Box\!\!\rightarrow (Z{=}z)\big) \to \vec{X}{=}\vec{x}, Y{=}y \,\Box\!\!\rightarrow (Z{=}z)$	
A_4	$\vec{X}{=}\vec{x}, Y{=}y \,\Box\!\!\rightarrow (Y{=}y)$	
A_5	$\big(\vec{X}{=}\vec{x}, Y{=}y \,\Box\!\!\rightarrow (Z{=}z) \wedge \vec{X}{=}\vec{x}, Z{=}z \,\Box\!\!\rightarrow (Y{=}y)\big) \to \vec{X}{=}\vec{x} \,\Box\!\!\rightarrow (Z{=}z)$	for $Y \neq Z$
A_6	$(X_0 \rightsquigarrow X_1 \wedge \cdots \wedge X_{k-1} \rightsquigarrow X_k) \to \neg(X_k \rightsquigarrow X_0)$	$X_0, ..., X_k$ are distinct variables
$A_\neg$	$\vec{X}{=}\vec{x} \,\Box\!\!\rightarrow \neg\varphi \leftrightarrow \neg\vec{X}{=}\vec{x} \,\Box\!\!\rightarrow \varphi$	
$A_\wedge$	$\vec{X}{=}\vec{x} \,\Box\!\!\rightarrow (\varphi_1 \wedge \varphi_2) \leftrightarrow (\vec{X}{=}\vec{x} \,\Box\!\!\rightarrow \varphi_1 \wedge \vec{X}{=}\vec{x} \,\Box\!\!\rightarrow \varphi_2)$	
$A_{[][]}$	$\vec{X}{=}\vec{x} \,\Box\!\!\rightarrow (\vec{Y}{=}\vec{y} \,\Box\!\!\rightarrow \varphi) \leftrightarrow \vec{X}'{=}\vec{x}', \vec{Y}{=}\vec{y} \,\Box\!\!\rightarrow \varphi$	with $\vec{X}'{=}\vec{x}'$ the subassignment of $\vec{X}{=}\vec{x}$ for $\vec{X}' := \vec{X} \setminus \vec{Y}$
A_7	$(\vec{V} = \vec{v} \,\Box\!\!\rightarrow \mathbb{D}(Y = y)) \leftrightarrow (\vec{V} = \vec{v} \,\Box\!\!\rightarrow (\vec{X} = \vec{x}) \,\Box\!\!\rightarrow \mathbb{D}(Y = y))$	with $Y \notin \vec{X}$ and $\vec{V} = \vec{v}$ is a full assignment to $\mathcal{V}$
A_8	$\vec{X}{=}\vec{x} \to \neg\mathbb{D}(\vec{X}{=}\vec{x})$	
A_9	$(\vec{X} = \vec{x} \,\Box\!\!\rightarrow Z = z) \to (\vec{X} = \vec{x} \,\Box\!\!\rightarrow \mathbb{D}(\vec{V} = \vec{v})) \leftrightarrow (\vec{X} = \vec{x}, Z = z \,\Box\!\!\rightarrow \mathbb{D}(\vec{V} = \vec{v}))$	if $\vec{V} = \vec{v}$ is a full assignment to $\mathcal{V}$ and $Z \notin \vec{X}$
A_{10}	$(\vec{X}{=}\vec{x} \,\Box\!\!\rightarrow \vec{Y}{=}\vec{y}) \to (\mathbb{D}(\vec{X}{=}\vec{x}) \leftrightarrow \mathbb{D}(\vec{X}{=}\vec{x}, \vec{Y}'{=}\vec{y}'))$	with $\vec{Y}' = \vec{y}'$ the sub-assignment of $\vec{Y}{=}\vec{y}$ for $\vec{Y}' := \vec{Y} \setminus \vec{X}$
A_{11}	$(\vec{X}{=}\vec{x} \,\Box\!\!\rightarrow D(\vec{Y}{=}\vec{y})) \wedge \mathbb{D}(\vec{X}{=}\vec{x}) \to \mathbb{D}(\vec{X}'{=}\vec{x}', \vec{Y}{=}\vec{y})$	with $\vec{X}' = \vec{x}'$ the sub-assignment of $\vec{X}{=}\vec{x}$ for $\vec{X}' := \vec{X} \setminus \vec{Y}$
A_{12}	$\vec{V} = \vec{v} \to (\mathbb{D}(\vec{V} = \vec{v}') \vee (\vec{V} = \vec{v}' \,\Box\!\!\rightarrow \mathbb{D}(\vec{V} = \vec{v})))$	$\vec{V} = \vec{v}$ and $\vec{V} = \vec{v}'$ are distinct full assignments to $\mathcal{V}$
A_{13}	$\neg(X < X)$	
A_{14}	$X < Y \vee Y < X$	if X and Y are distinct
A_{15}	$X < Y \wedge Y < Z \to X < Z$	
A_{16}	$\vec{Z} < Y \to (\vec{V} = \vec{v} \,\Box\!\!\rightarrow (\mathbb{D}(Y = y) \to \mathbb{D}(Y = y, \vec{Z} = \vec{z})))$	$\vec{V} = \vec{v}$ is a full assignment to $\mathcal{V}$
A_{17}	$Y < X \to (\vec{Z} = \vec{z} \,\Box\!\!\rightarrow Y < X)$	

Table 5.1 Axiom System L_{DC}

with the consequence of X is also desired. A_{11} indicates that if X is desired and Y is desired conditional on X, then $X \wedge Y$ must be desired as well (as far as they are disjoint). A_{12} says that any two situation (represented by full assignments) are comparable. A_{13}-A_{15} says that the priority order is strict and total. A_{16} says that if all the causal connection among variables is removed, then the desire of Y is independent of any change of variables which are less important than Y. A_{17} says that the priority order is invariant under any intervention.

Theorem 1: A_8-A_{17} *is sound with respect to desire causality models.*

Proof: See the Appendix A. □

Let $\mathbb{C}$ be the set of all maximally L_{DC}-consistent set of $\mathcal{L}_{DC}$-formulas. We will show that for any $\Gamma \in \mathbb{C}$, there is a DC-model $\mathcal{M}^\Gamma$ that models Γ.

Definition 43 (Building the canonical model)**:** *For any* $\Gamma \in \mathbb{C}$*, the canonical model* $\mathcal{M}^\Gamma = \langle \mathcal{V}, \mathcal{F}^\Gamma, \ll^\Gamma, In^\Gamma \rangle$ *is defined as below.*

- *For each variable $V \in \mathcal{V}$, the structural function $\mathcal{F} = \{f_V^\Gamma\}_{V \in \mathcal{V}}$ is defined in such a way that: for each possible assignment $\vec{X} = \vec{x}$ and any $v \in \mathcal{R}(V)$*

$$f_V^\Gamma(\vec{X} = \vec{x}) = v \quad \textit{if and only if} \quad \vec{X}{=}\vec{x} \;\Box\!\!\rightarrow V{=}v \in \Gamma$$

  Where $\vec{X}$ be a tuple with all variables in $\mathcal{V} \setminus V$ and $\vec{x}$ are digits in $\{0,1\}$.
- *$\ll^\Gamma$ is defined by $X \ll Y$ iff $X < Y \in \Gamma$*
- *In is defined as follows: $In(Z) = z$ iff for some $\vec{x}$, $\vec{X} = \vec{x}, Z = z' \;\Box\!\!\rightarrow \mathbb{D}(Z = z) \in \Gamma$ where $\vec{X}$ lists all the variables except Z and z' is different from z (we will show that In is well-defined later).*

A_1-A_6 guarantee that: (i) for each variable $V \in \mathcal{U} \cup \mathcal{V}$ and any assignment $\vec{X} = \vec{x}$ to $\mathcal{U} \cup \mathcal{V} \setminus V$, there is exactly one value $v \in \mathcal{R}(V)$ such that $\vec{X}{=}\vec{x} \;\Box\!\!\rightarrow V{=}v \in \Gamma$; (ii) for each variable $X \in \mathcal{U} \cup \mathcal{V}$ there is exactly one value of X such that $X = x \in \Gamma$. Therefore $\mathcal{F}^\Gamma$ and $\mathcal{A}^\Gamma$ is well defined.

A_7-A_{12} guarantee that In is well-defined. Suppose not, then for some $\vec{x}$, both $\vec{X}{=}\vec{x}, Z{=}z \;\Box\!\!\rightarrow \mathbb{D}(Z{=}z')$ and $\vec{X}{=}\vec{x}', Z{=}z' \;\Box\!\!\rightarrow \mathbb{D}(Z{=}z)$ are in Γ, with $z' \notin z$. By A_7, it follows that $\vec{X}{=}\vec{x}, Z{=}z' \;\Box\!\!\rightarrow \mathbb{D}(Z{=}z)$ is also in Γ. By generalization and A_{10}, we have $\vec{X}{=}\vec{x}, Z{=}z \;\Box\!\!\rightarrow \mathbb{D}(\vec{X}{=}\vec{x}, Z{=}z')$ and $\vec{X}{=}\vec{x}, Z{=}z' \;\Box\!\!\rightarrow \mathbb{D}(\vec{X}{=}\vec{x}, Z{=}z)$ both in Γ. By $A_{[][]}$, $\vec{X}{=}\vec{x}, Z{=}z \;\Box\!\!\rightarrow (\vec{X}{=}\vec{x}, Z{=}z' \;\Box\!\!\rightarrow \mathbb{D}(\vec{X}{=}\vec{x}, Z{=}z))$ is in Γ. By A_{11}, $\vec{X}{=}\vec{x}, Z{=}z \;\Box\!\!\rightarrow \mathbb{D}(\vec{X}{=}\vec{x}, Z{=}z)$. However Generalization and A_8 implies that $\vec{X}{=}\vec{x}, Z{=}z \;\Box\!\!\rightarrow \neg\mathbb{D}(\vec{X}{=}\vec{x}, Z{=}z)$. Contradiction. A_{12} guarantees that exactly one of $\vec{X}{=}\vec{x}, Z{=}z \;\Box\!\!\rightarrow \mathbb{D}(\vec{X}{=}\vec{x}, Z{=}z')$ and $\vec{X}{=}\vec{x}, Z{=}z' \;\Box\!\!\rightarrow \mathbb{D}(\vec{X}{=}\vec{x}, Z{=}z)$ are in Γ

A_{13}-A_{15} guarantee that $\ll^\Gamma$ is a strict total order.

Theorem 2 (Truth Lemma for the $\mathbb{D}$-free formulas in $\mathcal{L}_{DC}$)**:** $\vec{X} = \vec{x} \;\Box\!\!\rightarrow Y{=}y \in \Gamma$ *iff* $\langle \mathcal{V}, \mathcal{F}^\Gamma, \ll^\Gamma, In^\Gamma \rangle \vDash \vec{X}{=}\vec{x} \;\Box\!\!\rightarrow Y{=}y$

Proof: The truth lemma for the sub-language without desire operator has been proven in Halpern (2000): for any $\phi \in \mathcal{L}_C$ (the sub-language of $\mathcal{L}_{DC}$ without the $\mathbb{D}$ operator and $<$ operator), $\phi \in C$ iff $\langle \mathcal{V}, \mathcal{F}^C \vDash \phi \rangle$ where C is maximal consistent set with respect to L_C, where $\mathcal{F}^C$ is defined in the same way as in Definition 43. Notice that the truth condition of formulas does not involve preference structure, we can apply exactly the same strategy of Halpern (2000) and show that for any formula $\vec{X} = \vec{x} \;\Box\!\!\rightarrow Y = y \in \Gamma$ iff $\langle \mathcal{V}, \mathcal{F}^\Gamma_{\vec{X}=\vec{x}}, \ll^\Gamma, In^\Gamma \rangle \vDash Y{=}y$ iff $\langle \mathcal{V}, \mathcal{F}^\Gamma, \ll^\Gamma, In^\Gamma \rangle \vDash \vec{X}{=}\vec{x} \;\Box\!\!\rightarrow Y{=}y$

□

Theorem 3 (Truth Lemma for $\mathcal{L}_{DC}$)**:** *Let $\Gamma \in \mathbb{C}$, for any $\phi \in \mathcal{L}_{DC}$, $\phi \in \Gamma$ iff $\langle \mathcal{V}, \mathcal{F}^\Gamma, \ll^\Gamma$*

$, In^{\Gamma}\rangle \models \phi$

Proof: See the Appendix. □

Theorem 4: L_{DC} *is complete with respect to desire causality models.*

Proof: By Theorem 3, we can conclude that for any maximally L_{DC}-consistent set of $\mathcal{L}_{DC}$-formulas, there is a Desire-Causality model $\mathcal{M}^{\Gamma}$ that models Γ. □

5.5 Discussion

Conflicting desire According to our formalism of desire, Desire ϕ and Desire ψ does not imply Desire $\phi \wedge \psi$, namely $\mathbb{D}(X = x) \wedge \mathbb{D}(Y = y) \to \mathbb{D}(X = x, Y = y)$ is not valid. This feature fits our intuition that, in general, it is not the case that desiring one thing and desiring another means desiring the two things at the same time. Especially in situations where desired things contain conflict. Consider the following example of conflicting desires (cf. Levinson (2003)):

Example 3: *To have a happy holiday, Robin can and will only visit one European city this summer. Otherwise he'll get an unpleasant holiday because of a flustered and tight schedule. When making plans Robin wants to visit Paris and wants to visit Rome.*

Levinson (Levinson (2003)) called two desires conflicting if they cannot be fulfilled together. In the example, the desires of visiting Paris and visiting Rome are in conflict for Robin since the two desires exclude each other. Levinson argued that having contradicting desires is very common in daily life, so an adequate model of desire should explain this phenomenon. In his analysis, such conflicting desires can be captured in a probabilistic model.

However, predicting the conflicting desires to be felicitous is a challenge to the analyses that follow Hintikka (Hintikka (1969)) or Stalnaker (Stalnaker (1984)) traditions, such as Heim (1992); Von Fintel (1999); Kratzer (1981). In this chapter, inspired by Stalnaker's account of desire, we propose a logic L_{DC}, and the conflicting desires can be characterized by it.

In our analysis, going to Paris or going to Rome has a different effect than going to Rome and Paris. So we affirm Robin's desires of being Paris and being Rome because each visiting can lead to a happy holiday. But we reject his desire to visit both cities during the summer, which will cause a bad vacation.

To characterize this, we first distinguish $\mathbb{D}(X = x) \wedge \mathbb{D}(Y = y$ and $\mathbb{D}(X = x, Y = y)$

in $\mathcal{L}_{DC}$. We argue that the former indicates that the agent has two desires, while the latter expresses that she just has one desire where she wants to fulfill both propositions $X = x$ and $Y = y$. This distinction can be reflected by the following reasoning which holds in our system:

Fact 16: $\mathbb{D}(X = x) \wedge \mathbb{D}(Y = y) \nvDash \mathbb{D}(X = x, Y = y)$

This says that having desires for two propositions separately does not imply desire for the combination of these two propositions. Accordingly, in the holiday example, if Robin has desires of visiting Paris and visiting Rome separately, then we will not derive that he wants to visit both cities during the summer. In this way, we rule out the desire that is truly contradictory. As a result, visiting Paris or visiting Rome, as parts of the plan, can be desired separately in our account[1].

Independent desire based on causal graph The holiday example is a counterexample to $\mathbb{D}\phi \wedge \mathbb{D}\psi \to \mathbb{D}(\phi \wedge \psi)$, and the desire-causality model explains why $(\mathbb{D}(X = 1) \wedge \mathbb{D}(Y = 1)) \to \mathbb{D}(X = 1, Y = 1)$ is not valid according to our semantics.

However, there is still some natural connection between $\mathbb{D}\phi \wedge \mathbb{D}\psi$ and $\mathbb{D}(\phi \wedge \psi)$: suppose ϕ and ψ "have no causal relation", then it is plausible to say Desire $\phi \wedge \psi$ whenever Desire ϕ and Desire ψ.

This connection is reflected in desire-causality models. Let $\mathcal{M} = \langle \mathcal{V}, \mathcal{F}, \ll, In \rangle$ be a desire-causality model, if there is no variable that is a common descendent of X and Y, then $\mathcal{M} \vDash \mathbb{D}(X = x) \wedge \mathbb{D}(Y = y)$ if and only if $\mathcal{M} \vDash \mathbb{D}(X = x, Y = y)$.

This property can be expressed by $\mathcal{L}_{DC}$, that is: $(\wedge_{V \in \mathcal{V} \setminus \{X,Y\}} \neg(X \rightsquigarrow V \wedge Y \rightsquigarrow V)) \to (\mathbb{D}(X{=}x) \wedge \mathbb{D}(Y{=}y) \leftrightarrow \mathbb{D}(X{=}x, Y{=}y))$. This formula is valid according to L_{DC}.

The validity of $(\wedge_{V \in \mathcal{V} \setminus \{X,Y\}} \neg(X \rightsquigarrow V \wedge Y \rightsquigarrow V)) \to (\mathbb{D}(X{=}x) \wedge \mathbb{D}(Y{=}y) \leftrightarrow \mathbb{D}(X{=}x, Y{=}y))$ is an instance reflecting the entanglement between causal structure and preference structure in the desire-causality model.

Conditional desire Our language is also able to formulate conditional desire. It is intuitive to interpret conditional desire $\mathbb{D}(\phi|\psi)$ as "had ϕ been counterfactually forced to be true, then ϕ is desired ". Therefore "desire $\vec{Y} = \vec{y}$ conditional on $\vec{X} = \vec{x}$" can be formulated as $\vec{X} = \vec{x} \;\Box\!\!\to \mathbb{D}(\vec{Y} = \vec{y})$ in $\mathcal{L}_{DC}$.

Compared with the definition of conditional desire in Lang et al. (2003), our treatment has much in common with it if we regard atomic propositions as single variables. Lang et al. (2003) interprets $\mathbb{D}(Y|X)$ as: there are maximally normal $X \wedge Y$-worlds which are

1 See another interpretation of the consistency of this example in Lang et al. (2003).

strictly preferred to all the most normal $X \wedge \neg Y$-worlds. According to our approach, by Definition 42, the truth condition of "desiring $Y{=}1$ conditional on $X{=}1$" is equivalent to: $\mathcal{A}^{\mathcal{F}_{X=1}}$ is strictly more preferred than $\mathcal{A}^{\mathcal{F}_{X=1,Y=1}}$. As we have argued in section 3, we consider $\mathcal{A}^{\mathcal{F}_{X=1}}$ as the closest alternative where $X{=}1$ is forced to be true and $\mathcal{A}^{\mathcal{F}_{X=1,Y=1}}$ as the closest alternative where $X{=}1 \wedge Y{=}1$ is forced to be true. So the only difference with Lang et al. (2003) in the evaluation of $\mathbb{D}(Y|X)$ is that we compare the maximally normal $X \wedge Y$-worlds to the maximally normal X-worlds, instead of $X \wedge \neg Y$-worlds. For instance, if $X{=}1$ causally leads to $Y{=}1$, then according to our account, "desiring $Y{=}1$ conditional on $X{=}1$" does not hold.

According to our semantics, $((\phi \mathrel{\Box\!\!\rightarrow} \psi) \wedge (\phi \mathrel{\Box\!\!\rightarrow} \mathbb{D}\chi)) \rightarrow (\phi \wedge \psi \mathrel{\Box\!\!\rightarrow} \mathbb{D}\chi)$, namely $((\vec{X} = \vec{x} \mathrel{\Box\!\!\rightarrow} Z = z) \mathrel{\Box\!\!\rightarrow} (\vec{X} = \vec{x} \mathrel{\Box\!\!\rightarrow} \mathbb{D}(\vec{Y} = \vec{y}))) \rightarrow (\vec{X} = \vec{x}, Z = z \mathrel{\Box\!\!\rightarrow} \mathbb{D}(\vec{Y} = \vec{y}))$ is not valid (if $\vec{Y} = \vec{y}$ is not a full assignment to $\mathcal{V}$)[1]. To see this, consider the following example:

> *A rock is thrown on an expensive piece of glass. To protect the glass, Robin can block the rock with a board, or with his own body. The best scenario is obviously to both protect the glass and he himself is not injured. So Robin uses the board to block the rock.*

In the example, both 'using board' and 'using his own body' lead to 'the glass is not broken'. Robin has made a decision to use the board to block the rock. He can also have a conditional desire to use his own body, namely if Robin cannot find a board, he will want to block the rock by his body.

Let B, D and G stand for 'whether using a board to block the rock', 'whether using his body to block the rock' and 'Whether the glass is smashed' respectively. According to the example, the most important for Robin is to protect the glass and then protect himself. The least important thing is 'protect the board'. The priority of the variables is illustrated as follows:

$$B \ll D \ll G$$

In the example, Robin's intrinsic desire of the variables is $In(B) = 0$, $In(D) = 0$ and $In(G) = 0$. Namely, he primarily wants nothing to be hit by the rock. Accordingly, the derived preference of Robin can be concluded, which is described by Figure 5.3.

In the rock example, there is a counterfactual: B=0 $\Box\!\!\rightarrow$G=1. Robin decides to use a board to block the rock, so $\mathcal{A}_2$ is the actual assignment. $\mathcal{A}_5$ will be the assignment if the value of B is 0, and then $\mathcal{A}_3$ will be the assignment that results from forcing the value of D

1 Notice that we require $\vec{V} = \vec{v}$ to be a full assignment in axiom A_9.

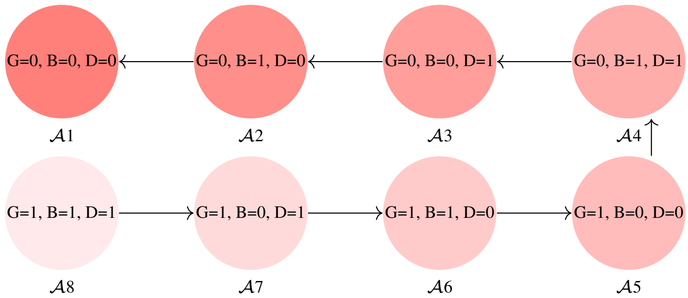


Figure 5.3 Derived preference of Robin

to be 1. As illustrated by Fig 5.3, $\mathcal{A}_5 \prec_{\mathcal{M}} \mathcal{A}_3$ holds. According to our semantics, we can verify that (B=0 $\Box\!\!\rightarrow$ ($\mathbb{D}$(D=1))). In addition, if we assume a condition that B=0 and G=1, then $\mathcal{A}_7$ will be the assignment after intervention forcing the value of D to be 1, rather than $\mathcal{A}_3$. The preference odering shows that $\mathcal{A}_7$ is not better that $\mathcal{A}_5$, that is $\mathcal{A}_7 \prec_{\mathcal{M}} \mathcal{A}_5$. Therefore we falsify $\mathbb{D}$(D=1) under the condition $B = 0, G = 1$. As a result, the formula ((B=0 $\Box\!\!\rightarrow$ G=1)$\Box\!\!\rightarrow$ (B=0 $\Box\!\!\rightarrow$ $\mathbb{D}$(D=1)))$\rightarrow$ (B=0, G=1 $\Box\!\!\rightarrow$ $\mathbb{D}$(D=1)) does not hold.

Intrinsic desire and extrinsic desire. As mentioned in Sec.5.3, the approach of modeling desire proposed in this chapter makes it possible to distinguish intrinsic desire and extrinsic desire. There is much discussion on the different notions of desire in philosophical literature, such as Arpaly and Schroeder (2014); Smith et al. (1989). In Liu (2011b), Liu proposes a two-level perspective on preference, in which the intrinsic preference is modeled based on the primitive betterness relations and the extrinsic preference is based on the priority structure as an evidence base for preference.

Likewise, in our proposal, both betterness relations and priority ordering make sense in distinguishing intrinsic desire and extrinsic desire. By desire-causality model defined in Def.39, the intrinsic desire can be captured by the function In. On the other hand, the causal structure in the model allows us to express the extrinsic desire which means something is desired for the outcomes it causes. Suppose a scenario is modeled by a desire-causality model $\langle \mathcal{V}, \mathcal{F}, \ll, In \rangle$, the agent has an intrinsic desire of $X = x$ whenever $In(X) = x$; the agent has an extrinsic desire of $X = x$ whenever $\langle \mathcal{V}, \mathcal{F}, \ll, In \rangle \models \mathbb{D}(X = x)$. In addition, taking advantage of causality combined preferences to characterize effect-oriented desires, we explore the relationship between both types of desires.

Intrinsic desire and extrinsic desire sometimes are compatible. For example, people want to be healthy, partially because of their preference of health itself, and partially because health can bring them beneficial effects such as enjoying life or work achievements. In this case, $D(X = x)$ is satisfied in our model as well as $In(X) = x$. However, there are also cases where something is desired only instrumentally, such like the situation described by the example of sleeping pills. Robin doesn't want to eat the pills intrinsically, but for the best effect (having a good sleep) he wants to do that. In our framework, such kind scenarios can be modeled by the cases where $\langle \mathcal{V}, \mathcal{F}, \ll, In \rangle \vDash \mathbb{D}(X = x)$ but $In(X) \neq x$. According to our formal system, we can prove the following property: when $In(X) \neq x$, $\langle \mathcal{V}, \mathcal{F}, \ll, In \rangle \vDash \mathbb{D}(X = x)$ if and only if, there is $Y \gg X$ with $In(Y) = y$, such that $\langle \mathcal{V}, \mathcal{F}, \ll, In \rangle \vDash X = x \,\Box\!\!\rightarrow Y = y$. It intuitively says that if ϕ is not desired intrinsically by an agent, then she has a desire of it if and only if there is a consequence of ϕ which is intrinsically desired and is more important than ϕ.

5.6 Summary

In this chapter, we proposed that causality plays an important role in characterizing desire and we provided a formal system that takes both preference perspective and causal perspective into account. In order to formalize our account, we combined the structural equation models in Pearl (2000); Halpern (2000); Ibeling and Icard (2020) and the preference structure in de Jongh and Liu (2009); Van Benthem et al. (2014). The preference structure characterizes the intrinsic desire of the agent and the priority order of the variables. And the structural equations provide the information about the causal effects of the variables, so that desire can be derived from it together with the underlying the preference structure. Such an account of desire explains our intuition in Robin's examples at the beginning. We developed a logic system based on these models and provide a complete axiomatization for it. At the end, in addition, we have shown that our system can be used to analyse other questions related to desire, such as the entanglement between desire and causal graph, the entanglement between intrinsic desire and extrinsic desire.

In this study, we do not examine the monotonicity of *want* based on monotonicity puzzles. Rather, within our logical framework, the monotonicity of *want* is indeed determined by causal relations and preference ordering, providing an alternative perspective on this semantic property.

CHAPTER 6 CONCLUSION

6.1 In retrospect

In this dissertation, we have undertaken an in-depth examination of modalities and monotonicity in modal context. The primary contribution of this thesis lies in the uniform account of WEM and investigation of the semantic and logical properties of modalities. We systematically address two key questions at both the implementation and framework levels.

WEM and reinterpretation theory. Through the examination of multiple puzzles of monotonicity under desire, we proposed that monotonicity reasoning, as a semantic principle, gives rise to pragmatic effects, such as free choice and ignorance effects. Free choice reasoning is normally triggered by disjunctive statements. We argued that sentences derived from upward monotonicity reasoning can be reinterpreted as disjunctive statements. By reinterpreting Q as $P \vee \neg P$, we can differentiate between the possibilities justified by the premises (P) and those that are not ($\neg P$). Reasoning about these unjustified possibilities is the underlying factor rendering monotonicity reasoning seemingly problematic. We implement this analysis under the BSML framework, as it offers a plausible formal account for such pragmatic reasoning. Consequently, we can perceive the WEM as a linguistic phenomenon bridging the boundary between language and logic as well as semantics and pragmatics.

Interaction between epistemic possibilies, knowledg and belief. Chapters 2, 3, and 5 of this thesis investigated specific modalities and concrete puzzles, under WEM theory. Chapters 2 and 3 delved into epistemic modalities, with a particular focus on the interaction among epistemic possibilities, knowledge, and belief. As evidenced by the puzzle presented in Chapter 2, the tension generated by monotonic effects results in so-called epistemic contradictions, which are also observed within the scope of the verb "believe". To address these issues, we expanded BSML into an epistemic framework, denoted BSEL. Within this framework, we examined the classical and non-classical properties of epistemic *might*, and subsequently explore the properties of knowledge and belief from the perspective of their interaction with epistemic *might*. Notably, we proposed a novel semantics of knowledge within BSEL, wherein the factivity of knowledge is treated as a presupposition rather than being solely defined by reflexivity.

Chapter 3 builds upon the foundation established in Chapter 2, further developing BSEL into a multi-agent framework. This chapter primarily concentrated on the perspective-sensitivity of epistemic *might*. By integrating the two-dimensional semantics of Epistemic Friendship Logic, we enabled perspective shifts. Within this framework, a broader range of interactions can be investigated.

Desire based on preference and causality. Chapter 5 was dedicated to examining the modality of desire. In this chapter, we argued that the semantics of desire are intimately connected to causality and preference. Consequently, we proposed a desire-causality model to interpret desire based on causal inferences. Furthermore, we investigated the logic of this model and provided an axiomatization. The conclusion drawn in Chapter 4 already shows that the WEM phenomenon does not influence the semantics of desire modals. As such, we can concentrate on the specific requirements for the modal semantics of desire itself.

In summary, this thesis shows the value of examining modalities through the lens of the WEM theory in addressing complex issues at the intersection of formal semantics, modal logic, and philosophy. Our investigation has spanned across various modal domains, including epistemic, desire also deontic domain. By developing novel frameworks, such as BSEL and the desire-causal logic, we have shed light on the underlying issues and connections within these domains. However, the research presented here serves as merely a starting point, with numerous intriguing questions and potential connections awaiting exploration in future work. The subsequent section will highlight some of these, building upon the discussion provided in the chapters of this thesis.

6.2 The road ahead

Throughout the core chapters of this thesis, namely Chapters 2 to 5, a new framework is introduced in almost every chapter. Consequently, numerous questions and potential developments present themselves, both from technical and conceptual standpoints. In this section, we shall propose a topic or work related to Chapter 2-5, which can serve as a foundation for continued exploration in future research.

Chapter 2. BSEL models provide a comprehensive characterization of the properties of epistemic modals, encompassing both classical and non-classical attributes. But we haven't studied its logical aspects and the axiomatization. Initially, we will study if the operator $\blacklozenge$ is definable in BSML. A possible way to define $\blacklozenge$ in BSML is that: $\blacklozenge\phi = (\phi \wedge \mathrm{NE}) \vee \top$. But this tranlsation does not capture the anit-support condition. Then, when incorporating plausibility ordering, it seems necessary to introduce additional axioms concerning beliefs. However, axiomatizing presupposition presents a more complex challenge. One potential solution could involve defining presupposition as conditional belief, although this approach may not fully capture its anti-support condition.

Chapter 3. Within the current multi-agent model, we employ a simplistic definitional method in which the cardinality of each information state is equivalent. Consequently, expressing maximally informed and maximally ignorant agents becomes challenging. Ideally, we aim to assign distinct sets of possible worlds to different agents, reflecting their unique information contexts. Under such assumptions, we can further investigate a broader range of knowledge notions, including distributed knowledge and common knowledge. One potential solution involves redefining the @ operator, characterizing it as transitions between information states rather than between agents. However, this requires us to refine the method of lifting evaluation parameters.

Chapter 4. The reinterpretation theory demonstrates potential applicability to various logical and linguistic issues. One such application involves extending the QUD framework by modifying the focus alternative set, as exemplified by Ju in Ju (2023). However, this application necessitates broadening the reinterpretation theory and its formalization to involve sentences beyond those that merely pertain to nouns or verbs. Future research will explore the reinterpretation theory in a more general sense, addressing this particular issue.

Chapter 5. Our logic L_{DC} has been proposed to model desire based on causal inference. But it is an unnatural restriction that $\mathcal{L}_{DC}$ only talks about desire for simple propositions, namely the desire operator D can be only applied to conjunction of atomic sentences, e.g. $D(\vec{X} = \vec{x})$. In everyday life, an agent's desires always appear in more complex forms: conjunction, disjunction or negation. It is meaningful to analyse the property of desire in complex forms. For instance, it is clear that "desiring not taking sleeping pills" implies "not desiring taking sleeping pills", but not the other way around. So to be able to represent the meaning of desire for complicated propositions, we need to extend the language $\mathcal{L}_{DC}$

such that it allows any Boolean combination of atomic formulas in the scope of D operator. One possible method of such extension is to introduce Briggs' *selection function* in Briggs (2012), which selects a set of possible worlds that satisfies complex propositions according to the causal information.

Connection to Natural Logic. As mentioned in Chapter 1, monotonicity, recognized as the cornerstone of Natural Logic (Van Benthem (1988, 1995); Valencia (1991); Moss (1987, 2015); Icard (2012) etc.), is notable for its rapidity and low polynomial complexity, especially when it comes to polarity computation. Nevertheless, the investigations in this dissertation have unveiled that, in various modal contexts, monotonicity reasoning triggers weakening which may lead to a pragmatic failure. This reflects the tension bewteen monotonicity and language use, and it reveals that monotonicity is not safe everywhere. Consequently, this observation calls for a thorough examination into the precise circumstances under which monotonicity inferences are deemed acceptable within the realm of natural language. Furthermore, it raises the question of whether we need higher complexity to determine their acceptability rather than mere polarity count. These findings may present a potential obstacle to the effectiveness of monotonicity calculus systems. As we look towards future research, it is essential that we delve deeper into the fundamental principles of WEM theory, with a particular emphasis on the rules of suspending. Concurrently, we aim to investigate the ramifications of these principles on the operation of monotonic calculus systems.

Natural Logic operates on the efficient and effective components of natural language reasoning, and it shows preference for employing smaller fragments of a larger system. While this dissertation delves into more complex and larger logics. It is attractive to inquire whether a smaller fragment and its axiomatization of these logics could address monotonicity in conjunction with modal operators. In this regard, a modal logic system related to natural logic, such as Icard and Moss (2023), could serve as an exemplary system to explore.

Taking into account the potential oversimplification of the complex interrelationships among monotonicity and modalities, the analyses presented in the chapters of this dissertation nonetheless underscore the validity of this approach as a promising path forward in addressing numerous unresolved issues. At a minimum, the analysis conducted within this thesis contributes insights to our understanding of modalities. In conclusion, this dissertation has laid the groundwork for future research to build upon, and it is our hope that the insights gained from our exploration of monotonicity and modalities will inspire further investigation in this fascinating area of language and logic, semantics and pragmatics.

路漫漫其修远兮, 吾将上下而求索。

The road ahead will be long, Our climb will be steep.

APPENDIX A SOUNDNESS AND COMPLETENESS OF L_{DC}

A.1 Soundness of L_{DC}

The soundness of A_1 to A_6 has been proven in Halpern (2000) (but in a slightly different form) and the soundness of $A_{[][]}$ is shown in Barbero et al. (2020). The soundness of $A_{\neg}$ and $A_{\wedge}$ is straightforward.

Theorem 1: A_7-A_{17} *is sound with respect to desire-causality models.*

Proof: Let $M = \langle \mathcal{V}, \mathcal{F}, \ll, In \rangle$

For A_7, suppose $M \models \vec{V} = \vec{v} \;\Box\!\!\rightarrow D(Y = y)$. Then the value of Y in $\vec{V} = \vec{v}$ must be y' that is different from y. Then $\vec{Z} = \vec{z}, Y = y' \prec_M \vec{Z} = \vec{z}, Y = y$ where $\vec{Z} = \vec{z}$ is the sub-assignment of $\vec{V} = \vec{v}$ to $\mathcal{V} \setminus Y$. It follows that $In(Y) = y$ (suppose not, then by the definition of $\prec_M$, $\vec{Z} = \vec{z}, Y = y \prec_M \vec{Z} = \vec{z}, Y = y'$, contradiction). It follows that $\vec{Z} = \vec{z}', Y = y' \prec_M \vec{Z} = \vec{z}', Y = y$ where $\vec{Z} = \vec{z}'$ is the assignment results from changing the value of $\vec{X}$ to $\vec{x}$ in $\vec{Z} = \vec{z}$. Since Y is not in $\vec{X}$, it follows that $M \models \vec{V} = \vec{v} \;\Box\!\!\rightarrow (\vec{X} = \vec{x} \;\Box\!\!\rightarrow D(Y = y))$.

We can check that by Definition 8, $\prec_M$ is a strict total order over $2^{\mathcal{V}}$. The soundness of A_8 follows immediately from the irreflexivity of $\prec_M$.

For the soundness of A_9, suppose $M \models \vec{X} = \vec{x} \;\Box\!\!\rightarrow Z = z$. Then $\mathcal{A}^{\mathcal{F}_{\vec{X}=\vec{x}}} = \mathcal{A}^{\mathcal{F}_{\vec{X}=\vec{x},Z=z}}$. Since $\vec{V} = \vec{v}$ is a full assignment, therefore $M \models \vec{X} = \vec{x} \;\Box\!\!\rightarrow \vec{V} = \vec{v}$ iff $\mathcal{A}^{\mathcal{F}_{\vec{X}=\vec{x}}} \prec_M \vec{V} = \vec{v}$ iff $\mathcal{A}^{\mathcal{F}_{\vec{X}=\vec{x},Z=z}} \prec_M \vec{V} = \vec{v}$ iff $M \models \vec{X} = \vec{x}, Z = z \;\Box\!\!\rightarrow \vec{V} = \vec{v}$.

For the soundness of A_{10}, suppose $M \models \vec{X} = \vec{x} \;\Box\!\!\rightarrow \vec{Y} = \vec{y}$, then $\mathcal{A}^{\mathcal{F}_{\vec{X}=\vec{x}}} = \mathcal{A}^{\mathcal{F}_{\vec{X}=\vec{x},\vec{Y}'=\vec{y}'}}$. Then $M \models D(\vec{X}=\vec{x})$ iff $\mathcal{A}^{\mathcal{F}} \prec_M \mathcal{A}^{\mathcal{F}_{\vec{X}=\vec{x}}}$ iff $\mathcal{A}^{\mathcal{F}} \prec_M \mathcal{A}^{\mathcal{F}_{\vec{X}=\vec{x},\vec{Y}'=\vec{y}'}}$ iff $M \models D(\vec{X}=\vec{x}, \vec{Y}'=\vec{y}')$.

The soundness of A_{11} and A_{12} follows from the transitivity and totality of $\prec_M$.

The soundness of A_{13} and A_{15} follows from the irreflexivity, transitivity and totality of $\ll$.

For the soundness of A_{16}, let $\mathcal{F}'$ be the set of structural functions in $M_{\vec{V}=\vec{v}}$. Suppose $\vec{Z} \ll Y$. As all functions $\mathcal{F}'$ are constant, $\vec{V} = \vec{v}$ and $\mathcal{A}^{\mathcal{F}'_{Y=y}}$ only differ in the value of Y. In addition, for all variables S with $Y \ll S$ $\mathcal{A}^{\mathcal{F}'_{\vec{Y}=\vec{y},\vec{Z}=\vec{z}}}(S) = \mathcal{A}^{\mathcal{F}'_{\vec{Y}=\vec{y}}}(S) = \mathcal{A}^{\mathcal{F}_{\vec{V}=\vec{v}}}(S)$. Therefore $\vec{V} = \vec{v} \prec_M \mathcal{A}^{\mathcal{F}'_{Y=y}}$ implies $\vec{V} = \vec{v} \prec_M \mathcal{A}^{\mathcal{F}'_{Y=y,\vec{Z}=\vec{z}}}$.

A_{17} is sound because In is invariant from M to $M_{\vec{Z}=\vec{z}}$.

□

A.2 Completeness of L_{DC}

By Definition 43 and Theorem 2, we can prove the following theorems and propositions:

Proposition 1: *Let $\vec{V}$ lists all the causal variables, $\mathcal{A}(\vec{V}) = \vec{v}$ and $\mathcal{A}'(\vec{V}) = \vec{v}$, $\vec{V} = \vec{v} \mathrel{\Box\!\!\rightarrow} D(\vec{V} = \vec{v}') \in \Gamma$ iff $\mathcal{A} \prec_{M^\Gamma} \mathcal{A}'$.*

Proof: (⇒) By A_{13} to A_{15}, $\vec{V}$ can be listed as $V_1, ..., V_n$ such that for any $0 < k \leq n-1$, $V_k < V_{k+1} \in \Gamma$. Since $\vec{V} = \vec{v} \mathrel{\Box\!\!\rightarrow} D(\vec{V} = \vec{v}') \in \Gamma$, by A_8, $\mathcal{A}$ must be different from $\mathcal{A}'$, namely there must be some V_k among $V_1, ..., V_n$ such that $\mathcal{A}(V_k) \neq \mathcal{A}(V_k)$. Let m be the greatest index that $\mathcal{A}(V_m) = v_m$, $\mathcal{A}'(V_m) = v'_m$ and $v_m \neq v'_m$. Without loss of generality, we assume $v_m = 0$ and $v'_m = 1$. Let $\vec{X}$ be all the variables except V_m, suppose for some $\vec{x}$, $\vec{X} = \vec{x}, V_m = 1 \mathrel{\Box\!\!\rightarrow} D(V_m = 0) \in \Gamma$, namely $In^\Gamma(V_m) = 0$. Then by A_{10}, $\vec{X} = \vec{x}, V_m = 1 \mathrel{\Box\!\!\rightarrow} D(\vec{X} = \vec{x}, V_m = 0) \in \Gamma$. Let $\vec{Z}$ be all the variables with $V_m < \vec{Z}$ and $\vec{Y}$ be all the variables with $\vec{Y} < V_m$, then $\vec{Z} = \vec{z}, \vec{Y} = \vec{y}, V_m = 1 \mathrel{\Box\!\!\rightarrow} D(\vec{Z} = \vec{z}, \vec{Y} = \vec{y}, V_m = 0) \in \Gamma$ where $\mathcal{A}(\vec{Z}) = \vec{z}$ and $\mathcal{A}(\vec{Y}) = \vec{y}$. Since $\vec{Y} < V_m$, by A_{16}, A_{17}, $\vec{Z} = \vec{z}, \vec{Y} = \vec{y}, V_m = 1 \mathrel{\Box\!\!\rightarrow} D(\vec{Z} = \vec{z}, \vec{Y} = \vec{y}', V_m = 0) \in \Gamma$ where $\mathcal{A}(\vec{Z}) = \vec{z}$ and $\mathcal{A}(\vec{Y}) = \vec{y}'$. This contradicts to $\vec{V} = \vec{v} \mathrel{\Box\!\!\rightarrow} D(\vec{V} = \vec{v}') \in \Gamma$. Therefore $In^\Gamma(V_m) = 1$. Since m is the greatest index such that $\mathcal{A}(V_m) \neq \mathcal{A}'(V_m)$, it follows that $\mathcal{A} \prec_{M^\Gamma} \mathcal{A}'$

(⇐) Suppose $\mathcal{A} \prec_{M^\Gamma} \mathcal{A}'$, then $\mathcal{A}$ can be written as $\vec{X} = \vec{x}, V_m = v_m, \vec{Y} = \vec{y}$ and $\mathcal{A}'$ can be written as $\vec{X} = \vec{x}, V_m = v'_m, \vec{Y} = \vec{y}'$ where $In(V_m) = v'_m$ and m is the greatest index such that $\mathcal{A}(V_m) \neq \mathcal{A}'(V_m)$. By definition, $In(V_m) = v'_m$ implies that there is some valuation of $\vec{x}''$ and $\vec{y}''$ such that $\vec{X} = \vec{x}'', V_m = v_m, \vec{Y} = \vec{y}'' \mathrel{\Box\!\!\rightarrow} D(V_m = v'_m) \in \Gamma$. By A_{10}, $\vec{X} = \vec{x}'', V_m = v_m, \vec{Y} = \vec{y}'' \mathrel{\Box\!\!\rightarrow} D(\vec{X} = \vec{x}'', V_m = v'_m, \vec{Y} = \vec{y}'') \in \Gamma$. By A_7, $\vec{X} = \vec{x}'', V_m = v_m, \vec{Y} = \vec{y}'' \mathrel{\Box\!\!\rightarrow} (\vec{X} = \vec{x}, \vec{Y} = \vec{y} \mathrel{\Box\!\!\rightarrow} D(\vec{X} = \vec{x}'', V_m = v'_m, \vec{Y} = \vec{y}'')) \in \Gamma$. By $A_{[][]}$ $\vec{X} = \vec{x}, V_m = v_m, \vec{Y} = \vec{y} \mathrel{\Box\!\!\rightarrow} D(\vec{X} = \vec{x}, V_m = v'_m, \vec{Y} = \vec{y}) \in \Gamma$. Since $\vec{Y} < V_n \in \Gamma$, by A_{16} $\vec{X} = \vec{x}, V_m = v_m, \vec{Y} = \vec{y} \mathrel{\Box\!\!\rightarrow} D(\vec{X} = \vec{x}, V_m = v'_m, \vec{Y} = \vec{y}') \in \Gamma$, which is equivalent to $\vec{V} = \vec{v} \mathrel{\Box\!\!\rightarrow} D(\vec{V} = \vec{v}') \in \Gamma$.

□

Theorem 3 (Truth Lemma for $\mathcal{L}_{DC}$): *Let $\Gamma \in \mathbb{C}$, for any $\phi \in \mathcal{L}_{DC}$, $\phi \in \Gamma$ iff $\langle \mathcal{V}, \mathcal{F}^\Gamma, \ll^\Gamma, In^\Gamma \rangle \models \phi$.*

Proof: For the simplicity of our notation we will write $\mathcal{F}^\Gamma, \ll^\Gamma, In^\Gamma$ as $\mathcal{F}, \ll, In$, and write the canonical model as $M = \langle \mathcal{V}, \mathcal{F}, \ll, In \rangle$

It is suffice to show that for any formula of the form $\vec{X} = \vec{x} \mathrel{\Box\!\!\rightarrow} D(\vec{Y} = \vec{y})$, $M \models \vec{X} = \vec{x} \mathrel{\Box\!\!\rightarrow} D(\vec{Y} = \vec{y})$ iff $\vec{X} = \vec{x} \mathrel{\Box\!\!\rightarrow} D(\vec{Y} = \vec{y}) \in \Gamma$. Notice that $D(\vec{Y} = \vec{y})$ is a special case of $\vec{X} = \vec{x} \mathrel{\Box\!\!\rightarrow} D(\vec{Y} = \vec{y})$ where $\vec{X} = \vec{x}$ is an empty sequence.

$M \models \vec{X} = \vec{x} \mathrel{\Box\!\!\rightarrow} D(\vec{Y} = \vec{y})$ holds whenever $\mathcal{A}^{\mathcal{F}_{\vec{X}=\vec{x}}} \prec_M \mathcal{A}^{\mathcal{F}_{\vec{X}=\vec{x'},\vec{Y}=\vec{y}}}$ where $\vec{X}'{=}\vec{x'}$ the sub-assignment of $\vec{X}{=}\vec{x}$ for $\vec{X}' := \vec{X} \setminus \vec{Y}$. Assume $\mathcal{A}^{\mathcal{F}_{\vec{X}=\vec{x}}}(\vec{V}) = \vec{v}$ and $\mathcal{A}^{\mathcal{F}_{\vec{X}=\vec{x'},\vec{Y}=\vec{y}}}(\vec{V}) = \vec{v'}$ where $\vec{V}$ lists all the causal variables. Then $M_{\vec{X}=\vec{x}} \models \vec{V} = \vec{v}$ and $(M_{\vec{X}=\vec{x}})_{\vec{Y}=\vec{y}} \models \vec{V} = \vec{v'}$. Let $\vec{Z} = \vec{z'}$ be the sub-assignment of $\vec{V} = \vec{v'}$ to all the variables other than $\vec{X}$ and $\vec{Y}$. By Theorem 2, $\vec{X} = \vec{x} \mathrel{\Box\!\!\rightarrow} \vec{V} = \vec{v} \in \Gamma$ and $\vec{X} = \vec{x'}, \vec{Y} = \vec{y} \mathrel{\Box\!\!\rightarrow} \vec{Z} = \vec{z'} \in \Gamma$. Then we have:
$\mathcal{A}^{\mathcal{F}_{\vec{X}=\vec{x}}} \prec_M \mathcal{A}^{\mathcal{F}_{\vec{X}=\vec{x'},\vec{Y}=\vec{y}}}$
iff $\vec{V} = \vec{v} \mathrel{\Box\!\!\rightarrow} D(\vec{X} = \vec{x'}, \vec{Y} = \vec{y}, \vec{Z} = \vec{z}) \in \Gamma$ (By Proposition 1)
iff $\vec{X} = \vec{x} \mathrel{\Box\!\!\rightarrow} D(\vec{X} = \vec{x'}, \vec{Y} = \vec{y}, \vec{Z} = \vec{z})$ (by A_9 and $\vec{X} = \vec{x} \mathrel{\Box\!\!\rightarrow} \vec{V} = \vec{v} \in \Gamma$)
iff $\vec{X} = \vec{x} \mathrel{\Box\!\!\rightarrow} D(\vec{X} = \vec{x'}, \vec{Y} = \vec{y}) \in \Gamma$ (By A_{10} and $\vec{X} = \vec{x} \mathrel{\Box\!\!\rightarrow} \vec{Z} = \vec{z} \in \Gamma$)
iff $\vec{X} = \vec{x} \mathrel{\Box\!\!\rightarrow} D(\vec{Y} = \vec{y}) \in \Gamma$ (As $\vec{X} = \vec{x} \mathrel{\Box\!\!\rightarrow} \vec{X}' = \vec{x'} \in \Gamma$).

□

Theorem 4: *L_{DC} is complete with respect to desire causality models.*

Proof: By Theorem 3, we can conclude that for any maximally L_{DC}-consistent set of $\mathcal{L}_{DC}$-formulas, there is a desire-causality model M^Γ that models Γ.

□

Proposition 2: *Let X and Y be two distinct variables, $(\wedge_{V \in \mathcal{V} \setminus \{X,Y\}} \neg(X \rightsquigarrow V \wedge Y \rightsquigarrow V)) \rightarrow (D(X{=}x) \wedge D(Y{=}y) \leftrightarrow D(X{=}x, Y{=}y))$ is valid in L_{DC}.*

Proof: Let $M = \langle \mathcal{V}, \mathcal{F}, \ll, In \rangle$. Let $\vec{A}$ be the variables in $\{V \in \mathcal{V} | X{=}V \text{ or } X \rightsquigarrow V\}$ and $\vec{B}$ be the variables in $\{V \in \mathcal{V} | Y{=}V \text{ or } Y \rightsquigarrow V\}$. Suppose $M \models \wedge_{V \in \mathcal{V} \setminus \{X,Y\}} \neg(X \rightsquigarrow V \wedge Y \rightsquigarrow V)$, then $\vec{A}$ and $\vec{B}$ are disjoint. Therefore we can write $\mathcal{A}^{\mathcal{F}}$ as $\vec{A} = \vec{a}, \vec{B} = \vec{b}, \vec{Z} = \vec{z}$. Since X has no causal influence on $\vec{B}$ and Y has no causal influence on $\vec{A}$, $\mathcal{A}^{\mathcal{F}_{X=x}}$ can be written as $\vec{A} = \vec{a'}, \vec{B} = \vec{b}, \vec{Z} = \vec{z}$, and $\mathcal{A}^{\mathcal{F}_{Y=y}}$ can be written as $\vec{A} = \vec{a}, \vec{B} = \vec{b'}, \vec{Z} = \vec{z}$. Suppose $M \models D(X = x) \wedge D(Y = y)$. By definition of $\prec_M$, $\mathcal{A}^{\mathcal{F}} \prec_M \mathcal{A}^{\mathcal{F}_{X=x}}$ and $\mathcal{A}^{\mathcal{F}} \prec_M \mathcal{A}^{\mathcal{F}_{Y=y}}$ whenever: (i) there is some $B_j \in \vec{B}$ such that $In(A_i) = a_i$ and A_i is prior to any A_k with $In(A_k) \neq a'_k$; (ii) there is some $B_j \in \vec{B}$ such that $In(B_j) = b'_j$ and B_j is prior to any B_k with $In(B_k) \neq b'_k$. Without loss of generality, we assume $A_i \ll B_j$. Then for any variable that is not in $\vec{Z}$: if it is some A_k among $\vec{A}$, then $In(A_k) \neq a'_k$ implies $A_k \ll A_i \ll B_j$; if it is some B_k among $\vec{B}$, then $In(B_k) \neq b'_k$ implies $B_k \ll B_j$. Therefore $\mathcal{A}^{\mathcal{F}} \prec_M \mathcal{A}^{\mathcal{F}_{X=x}}$ and $\mathcal{A}^{\mathcal{F}} \prec_M \mathcal{A}^{\mathcal{F}_{X=x}}$ iff $\vec{A} = \vec{A}, \vec{B} = \vec{B}, \vec{Z} = \vec{z} \prec_M \vec{A} = \vec{a'}, \vec{B} = \vec{b'}, \vec{Z} = \vec{z}$ iff $\mathcal{A}^{\mathcal{F}} \prec \mathcal{A}^{\mathcal{F}_{X=x,Y=y}}$ iff

$M \models D(X = x, Y = y)$.

□

BIBLIOGRAPHY

Aloni, M. (2001). *Quantification under Conceptual Covers*. PhD thesis, ILLC, University of Amsterdam.

Aloni, M. (2007). Free choice, modals, and imperatives. *Natural Language Semantics*, 15(1):65–94.

Aloni, M. (2022). Logic and conversation: the case of free choice. *Semantics and Pragmatics*, 15:5.

Aloni, M. (2023). Neglect-zero effects in dynamic semantics. In Deng, D., Liu, M., Westerståhl, D., and Xie, K., editors, *The Third Tsinghua Interdisciplinary Workshop on Logic, Language, and Meaning: Dynamics in Logic and Language*, volume 13524 of *Lecture Notes in Computer Science*. Springer.

Aloni, M. and van Ormondt, P. (2023). Modified numerals and split disjunction: the first-order case. *Journal of Logic, Language and Information*. To appear.

Alonso-Ovalle, L. (2006). *Disjunction in alternative semantics*. PhD thesis, UMass Amherst.

Anttila, A. (2021). The logic of free choice. axiomatizations of state-based modal logics. Master's thesis, ILLC, University of Amsterdam.

Aqvist, L. (1967). Good samaritans, contrary-to-duty imperatives, and epistemic obligations. *Noûs*, 1(4):361–379.

Areces, C. and ten Cate, B. (2007). Hybrid logics. In Blackburn, P., Van Benthem, J., and Wolter, F., editors, *Handbook of Modal Logic*, volume 3 of *Studies in Logic and Practical Reasoning*, pages 821–868. Elsevier.

Armstrong, D. M. (1973). *Belief, truth and knowledge*. Cambridge University Press.

Arpaly, N. and Schroeder, T. (2014). *In praise of desire*. Oxford University Press.

Asher, N. (1987). A typology for attitude verbs and their anaphoric properties. *Linguistics and Philosophy*, 10(2):125–197.

Baltag, A., Renne, B., and Smets, S. (2014). The logic of justified belief, explicit knowledge, and conclusive evidence. *Annals of Pure and Applied Logic*, 165(1):49–81.

Baltag, A. and Smets, S. (2016). A qualitative theory of dynamic interactive belief revision. In Arló-Costa, H., Hendricks, V., and van Benthem, J., editors, *Readings in Formal Epistemology*, volume 1 of *Springer Graduate Texts in Philosophy*. Springer.

Barbero, F., Schulz, K., Smets, S., Velázquez-Quesada, F. R., and Xie, K. (2020). Thinking about causation: A causal language with epistemic operators. In *International Workshop on Dynamic Logic*, pages 17–32. Springer.

Barwise, J. and Cooper, R. (1981). Generalized quantifiers and natural language. *Linguistics and Philosophy*, 4(2):159–219.

Beddor, B. and Goldstein, S. (2018). Believing epistemic contradictions. *The Review of Symbolic Logic*, 11(1):87–114.

Blackburn, P. and Seligman, J. (1995). Hybrid languages. *Journal of Logic, Language and Information*, 4:251–272.

Boutilier, C. (1994). Conditional logics of normality: a modal approach. *Artificial intelligence*, 68(1):87–154.

Bricker, A. M. (2018). *Visuomotor noise and the non-factive analysis of knowledge*. PhD thesis, University of Edinburgh.

Briggs, R. (2012). Interventionist counterfactuals. *Philosophical studies*, 160(1):139–166.

Brooks, L. R. (1987). Decentralized control of categorization: The role of prior processing episodes. In Neisser, U., editor, *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization*, pages 141–174. Cambridge University Press.

Buckwalter, W. and Turri, J. (2020). Knowledge and truth: a skeptical challenge. *Pacific Philosophical Quarterly*, 101(1):93–101.

Büring, D. (2008). The least at least can do. In *West Coast Conference on Formal Linguistics (WCCFL)*, volume 26, pages 114–120.

Cariani, F. (2013). ‘Ought’ and resolution semantics. *Noûs*, 47(3):534–558.

Castañeda, H. (1981). The paradoxes of deontic logic: The simplest solution to all of them in one fell swoop. In Hilpinen, R., editor, *New Studies in Deontic Logic*, volume 152 of *Synthese Library*. Springer.

Chemla, E. (2009). Universal implicatures and free choice effects: experimental data. *Semantics and Pragmatics; Vol 2 (2009); 2:1-33*, 2.

Coste-Marquis, S., Lang, J., Liberatore, P., and Marquis, P. (2004). Expressive power and succinctness of propositional languages for preference representation. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Ninth International Conference (KR2004)*, pages 203–212. AAAI Press.

Counihan, M. (2008). *Looking for logic in all the wrong places: An investigation of language, literacy and logic in reasoning*. PhD thesis, ILLC, University of Amsterdam.

Crnič, L., Chemla, E., and Fox, D. (2015). Scalar implicatures of embedded disjunction. *Natural Language Semantics*, 23(4):271–305.

Crnič, L. (2011). *Getting even*. PhD thesis, Massachusetts Institute of Technology.

Davidson, D. (2001). Epistemology and truth. In *Subjective, Intersubjective, Objective: Philosophical Essays Volume 3*. Oxford University Press.

de Jongh, D. and Liu, F. (2009). Preference, priorities and belief. In *Preference Change: Approaches from Philosophy, Economics and Psychology*, pages 85–108. Theory and Decision Library.

Dekker, P. (1993). *Transentential Meditations: Ups and Downs in Dynamic Semantics*. PhD thesis, ILLC, University of Amsterdam.

DeRose, K. (1991). Epistemic possibilities. *Philosophical Review*, 100(4):581–605.

Egan, A., Hawthorne, J., and Weatherson, B. (2005). Epistemic modals in context. In Preyer, G. and Peter, G., editors, *Contextualism in Philosophy: Knowledge, Meaning, and Truth*. Clarendon Press.

Eijck, J. (2005). Syllogistics = monotonicity + symmetry + existential import. *Report - Software Engineering*, 1:1–15.

Fauconnier, G. (1975). Polarity and the scale principle. *Chicago Linguistics Society*, 11:188–190.

Fauconnier, G. (1979). Implication reversal in a natural language. In Guenthner, F. and Schmidt, S. J., editors, *Formal Semantics and Pragmatics for Natural Languages*, pages 289–302. Reidel.

Forrester, J. W. (1984). Gentle murder, or the adverbial samaritan. *The Journal of Philosophy*, 81(4):193–197.

Fox, C. (2010). The good samaritan and the hygienic cook: A cautionary tale about linguistic data. In Stalmaszczyk, P., editor, *Objects of Inquiry in Philosophy of Language and Linguistics*, page 103. Ontos Verlag.

Fox, C. (2012). Obligations and permissions. *Language and Linguistics Compass*, 6(9):593–610.

Fox, D. (2007). Free choice and the theory of scalar implicatures. In Sauerland, U. and Stateva, P., editors, *Presupposition and Implicature in Compositional Semantics*, pages 71–120. Palgrave Macmillan.

Fraassen, B. C. V. (1972). The logic of conditional obligation. *Journal of Philosophical Logic*, 1(3/4):417.

Galles, D. and Pearl, J. (1998). An axiomatic characterisation of causal counterfactuals. *Foundations of Science*, 1:151–182.

Gazdar, G. (1979). Pragmatics, implicature, presuposition and logical form. *Critica*, 12(35):113–122.

Geach, P. T. (1956). Good and evil. *Analysis*, 17(2):33–42.

Geurts, B. (1998). Presuppositions and anaphors in attitude contexts. *Linguistics and philosophy*, 21(6):545–601.

Geurts, B. (2005). Entertaining alternatives: Disjunctions as modals. *Natural Language Semantics*, 13(4):383–410.

Gillies, A. S. (2001). A new solution to moore's paradox. *Philosophical Studies*, 105(3):237–250.

Hacquard, V. and Wellwood, A. (2012). Embedding epistemic modals in english: a corpus-based study. *Semantics and Pragmatics*, 5(4):1–29.

Halpern, J. Y. (1997). Defining relative likelihood in partially-ordered preferential structures. *Journal of Artificial Intelligence Research*, 7:1–24.

Halpern, J. Y. (2000). Axiomatizing causal reasoning. *Journal of Artificial Intelligence Research*, 12:317–337.

Halpern, J. Y. (2013). From causal models to counterfactual structures. *The Review of Symbolic Logic*, 6(2):305–322.

Halpern, J. Y. (2017). *Reasoning about uncertainty*. MIT press.

Haslinger, N. (2020). Quantificational DP arguments of opaque predicates in german. Master's thesis, University of Vienna.

Hawke, P. and Steinert-Threlkeld, S. (2021). Semantic expressivism for epistemic modals. *Linguistics and Philosophy*, 44(2):475–511.

Hawthorne, J., Rothschild, D., and Spectre, L. (2016). Belief is weak. *Philosophical Studies*, 173(5):1393–1404.

Hazlett, A. (2010). The myth of factive verbs. *Philosophy and Phenomenological Research*, 80(3):497–522.

Heim, I. (1983). On the projection problem for presuppositions. In Portner, P. and Partee, B. H., editors, *Formal Semantics - the Essential Readings*, pages 249–260. Blackwell.

Heim, I. (1992). Presupposition projection and the semantics of attitude verbs. *Journal of semantics*, 9(3):183–221.

Hintikka, J. (1962). *Knowledge and belief: an introduction to the logic of the two notions*. Cornell University Press.

Hintikka, J. (1969). Semantics for prepositional attitudes. In Davis, J. W. et al., editors, *Philosophical Logic*, pages 21–45. Reidel.

Holliday, W. and Mandelkern, M. (2022). Compatibility, compossibility, and epistemic modality. In Degano, M., Roberts, T., Sbardolini, G., and Schouwstra, M., editors, *Proceedings of the 23Rd Amsterdam Colloquium*.

Horn, L. R. (1978). Remarks on neg-raising. In *Pragmatics*, pages 129–220. Brill.

Horty, J. (2007). Reasons as defaults. *Philosophers' Imprint*, 7:1–28.

Ibeling, D. and Icard, T. (2020). On open-universe causal reasoning. In Adams, R. P. and Gogate, V., editors, *Proceedings of The 35th Uncertainty in Artificial Intelligence Conference*, volume 115 of *Proceedings of Machine Learning Research*, pages 1233–1243.

Icard, T. (2012). Inclusion and exclusion in natural language. *Studia Logica*, 100(4):705–725.

Icard, T. and Moss, L. (2023). A simple logic of concepts. *Journal of Philosophical Logic*, 52:705–730.

Jackson, F. and Pargetter, R. (1986). Oughts, options, and actualism. *Philosophical Review*, 95(2):233–255.

Jerzak, E. (2019). Two ways to want? *The Journal of Philosophy*, 116(2):65–98.

Ju, C. (2023). *The expression and semantic inference of factivity*. PhD thesis, Peking University.

Kamp, H. (1974). Free choice permission. *Proceedings of the Aristotelian Society*, 74(1):57–74.

Karttunen, L. (1973). Presuppositions of compound sentences. *Linguistic Inquiry*, 4(2):167–193.

Karttunen, L. (1974). Presupposition and linguistic context. *Theoretical Linguistics*, 1:181–194.

Keenan, D. and Westerståhl, D. (1996). Generalized quantifiers in linguistics and logic. In Van Benthem, J. and Ter Meulen, A., editors, *Handbook of Logic and Language*, pages 837–893. Elsevier.

Kelp, C. and Pritchard, D. (2009). Two Deflationary Approaches to Fitch-Style Reasoning. In *New Essays on the Knowability Paradox*. Oxford University Press.

Kontinen, J. and Väänänen, J. (2009). On definability in dependence logic. *Journal of Logic, Language and Information*, 18(3):317–332.

Kratzer, A. (1977). What 'must' and 'can' must and can mean. *Linguistics and Philosophy*, 1(3):337–355.

Kratzer, A. (1981). The notional category of modality. In Eikmeyer, H. J. and Rieser, H., editors, *Words, Worlds, and Contexts: New Approaches in Word Semantics*, pages 38–74. De Gruyter.

Kratzer, A. (1986). Conditionals. *Chicago Linguistics Society*, 22(2):1–15.

Kratzer, A. (1991). Modality. In von Stechow, A. and Wunderlich, D., editors, *Semantics: An International Handbook of Contemporary Research*, pages 639–650. De Gruyter.

Kripke, S. A. (1963). Semantical analysis of modal logic I normal modal propositional calculi. *Mathematical Logic Quarterly*, 9(5-6):67–96.

Ladusaw, W. (1979). *Polarity sensitivity as inherent scope relations*. PhD thesis, University of Texas at Austin.

Lang, J. (1996). Conditional desires and utilities-an alternative approach to qualitative decision theory. In *Proceedings of the European Conference on Artificial Intelligence*, pages 318–322.

Lang, J., Van Der Torre, L., and Weydert, E. (2003). Hidden uncertainty in the logical representation of desires. In *IJCAI*, pages 685–690.

Lasersohn, P. (2005). Context dependence, disagreement, and predicates of personal taste. *Linguistics and Philosophy*, 28(6):643–686.

Lasersohn, P. (2009). Relative truth, speaker commitment, and control of implicit arguments. *Synthese*, 166(2):359–374.

Lassiter, D. (2011). *Measurement and modality: The scalar basis of modal semantics*. PhD thesis, New York University.

Levinson, D. (2003). Probabilistic model-theoretic semantics for'want'. In *Semantics and Linguistic Theory*, volume 13, pages 222–239.

Lewis, D. (1973). Counterfactuals and comparative possibility. In *Ifs*, pages 57–85. Springer.

Lewis, D. (1979). A problem about permission. In Saarinen, E., Hilpinen, R., Niiniluoto, I., and Hintikka, M., editors, *Essays in Honour of Jaakko Hintikka*, volume 124 of *Synthese Library*. Springer.

Liu, F. (2007). *Changing for the better: Preference dynamics and agent diversity*. PhD thesis, ILLC, University of Amsterdam.

Liu, F. (2011a). *Reasoning about preference dynamics*. Synthese Library. Springer Science & Business Media.

Liu, F. (2011b). A two-level perspective on preference. *Journal of Philosophical Logic*, 40(3):421–439.

Liu, F., Seligman, J., and Girard, P. (2014). Logical dynamics of belief change in the community. *Synthese*, 191(11):2403–2431.

Ludlow, P. (2005). Contextualism and the new linguistic turn in epistemology. In Preyer, G. and Peter, G., editors, *Contextualism in Philosophy: Knowledge, Meaning, and Truth*, pages 11–51. Oxford University Press.

MacCartney, B. (2009). *Natural language inference*. Stanford University press.

MacFarlane, J. (2011). Epistemic modals are assessment-sensitive. In Egan, A. and Weatherson, B., editors, *Epistemic Modality*. Oxford University Press.

MacFarlane, J. (2014). *Assessment Sensitivity: Relative Truth and its Applications*. Oxford University Press.

Mandelkern, M. (2019). Bounded modality. *Philosophical Review*, 128(1):1–61.

Marty, P., Romoli, J., Sudo, Y., and Breheny, R. (2023). What makes an inference robust? lingbuzz/006205.

McCarthy, J. J. (2007). What is optimality theory? *Language and Linguistics Compass*, 1(4):260–291.

McNamara, P. and Van De Putte, F. (2022). Deontic Logic. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Spring 2022 edition.

Mendia, J. A. (2015). Conveying ignorance ignorance inferences with superlative numeral modifiers. *Proceedings of ConSOLE XXIII*, 150:174.

Meyer, M. C. (2020). Free choice disjunction. In Gutzmann, D., Matthewson, L., Meier, C., Rullmann, H., and Zimmermann, T., editors, *The Wiley Blackwell Companion to Semantics*. Wiley Blackwell.

Mitova, V. (2018). *The Factive Turn in Epistemology*. Cambridge University Press.

Moss, L. S. (1987). Boolean semantics for natural language. *Journal of Symbolic Logic*, 52(2):554–555.

Moss, L. S. (2015). Natural logic. In Lappin, S. and Fox, C., editors, *Handbook of Contemporary Semantic Theory*. Wiley-Blackwell.

Myers-Schulz, B. and Schwitzgebel, E. (2013). Knowing that p without believing that p. *Noûs*, 47(2):371–384.

Ninan, D. (2005). Two puzzles about deontic necessity. In Gajewski, J., Hacquard, V., Nickel, B., and Yalcin, S., editors, *New Work on Modality*, volume 51 of *MIT Working Papers in Linguistics*, pages 149–178. MIT Working Papers in Linguistics.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, 115(1):39–57.

Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika*, 82(4):669–688.

Pearl, J. (2000). *Causality. Models, reasoning, and inference*. Cambridge University Press.

Peters, S. and Westerstahl, D. (2006). *Quantifiers in language and logic*. Clarendon Press.

Portner, P. (2009). *Modality*. Oxford Surveys in Semantics and Pragmatics. Oxford University Press.

Prince, A. and Smolensky, P. (2008). *Optimality theory: constraint interaction in generative grammar*, pages 1 – 71. Rutgers University Center for Cognitive Science.

Prior, A. N. (1958). Escapism: The logical basis of ethics. In Melden, A. I., editor, *Journal of Symbolic Logic*, pages 610–611. University of Washington Press.

Radford, C. (1966). Knowledge—by examples. *Analysis*, 27(1):1–11.

Ramotowska, S., Marty, P., Romoli, J., Sudo, Y., and Breheny, R. (2022). Diversity with universality. In Degano, M., Roberts, T., Sbardolini, G., and Schouwstra, M., editors, *Proceedings of the 23Rd Amsterdam Colloquium*.

Roberts, C. (1996). Information structure in discourse: towards an integrated formal theory of pragmatics. In *Ohio State University Working Papers in Linguistics*, volume 49. Ohio State University.

Rosch, E. (1978). Principles of categorization. In *Cognition and categorization*, pages 27–48. Lawrence Erlbaum Associates.

Ross, A. (1944). Imperatives and logic. *Philosophy of Science*, 11(1):30–46.

Ruan, J. and Thielscher, M. (2011). A logic for knowledge flow in social networks. In *AI 2011: Advances in Artificial Intelligence: 24th Australasian Joint Conference, Perth, Australia, December 5-8, 2011. Proceedings 24*, pages 511–520. Springer.

Sayre-McCord, G. (1986). Deontic logic and the priority of moral theory. *Noûs*, 20(2):179.

Schroeder, M. (2008). Expression for expressivists. *Philosophy and Phenomenological Research*, 76(1):86.

Seligman, J. (1997). The logic of correct description. *Advances in intensional logic*, pages 107–135.

Seligman, J., Liu, F., and Girard, P. (2011). Logic in the community. In Banerjee, M. and Seth, A., editors, *Logic and Its Applications*, pages 178–188. Springer Berlin Heidelberg.

Seligman, J., Liu, F., and Girard, P. (2013). Facebook and the epistemic logic of friendship. In *Poster presentation at TARK 2013 (arXiv preprint arXiv:1310.6440).*

Shi, C., Smets, S., and Velázquez-Quesada, F. R. (2018). Beliefs supported by binary arguments. *Journal of Applied Non-Classical Logics*, 28(2-3):165–188.

Shi, C. and Sun, Y. (2021). Logic of convex order. *Studia Logica*, 109(5):1019–1047.

Simons, M. (2005). Dividing things up: The semantics of or and the modal/or interaction. *Natural Language Semantics*, 13(3):271–316.

Smith, M., Lewis, D., and Johnston, M. (1989). Dispositional theories of value. In *Aristotelian Society Supplementary Volume*, volume 63, pages 89–174. Oxford University Press.

Sorensen, R. A. (1988). *Blindspots*. Oxford, England: Oxford University Press.

Speas, P. and Tenny, C. (2003). Configurational properties of point of view roles. *Asymmetry in grammar*, pages 315–344.

Spector, B. and Egré, P. (2015). A uniform semantics for embedded interrogatives: an answer, not necessarily the answer. *Synthese*, 192:1729–1784.

Spohn, W. (1988). Ordinal conditional functions. a dynamic theory of epistemic states. In Harper, W. L. and Skyrms, B., editors, *Causation in Decision, Belief Change, and Statistics, vol. II*. Kluwer Academic Publishers.

Stalnaker, R. (1968). A theory of conditionals. In *Ifs*, pages 41–55. Springer.

Stalnaker, R. (1984). *Inquiry*. Cambridge University Press.

Stalnaker, R. (2006). On logics of knowledge and belief. *Philosophical Studies*, 128(1):169–199.

Stenning, K. and van Lambalgen, M. (2008). *Human Reasoning and Cognitive Science*. MIT Press.

Stephenson, T. (2007). Judge dependence, epistemic modals, and predicates of personal taste. *Linguistics and Philosophy*, 30(4):487–525.

Stojanovic, I. (2007). Talking about taste: disagreement, implicit arguments, and relative truth. *Linguistics and Philosophy*, 30(6):691–706.

Sun, Z. and Liu, F. (2021). The inference pattern mou in mohist logic: A monotonicity reasoning view. *Roczniki Filozoficzne*, 68(4):257–270.

Thibaut, J.-P. and Gelaes, S. (2006). Exemplar effects in the context of a categorization rule: Featural and holistic influences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(6):1403.

Tomberlin, J. E. and Mcguinness, F. (1977). "Because" and good samaritans. *Crítica: Revista Hispanoamericana de Filosofía*, 9(27):67–81.

Turri, J. (2011). Mythology of the factive. *Logos and Episteme*, 2(1):143–152.

Väänänen, J. (2007). Team logic. In *Interactive logic. Selected papers from the 7th Augustus de Morgan workshop*, pages 281–302. Amsterdam University Press.

Vainikka, A. (1987). Why can or mean and or or. *University of Massachusetts Occasional Papers in Linguistics*, 13.

Valencia, V. M. S. (1991). *Studies on natural logic and categorial grammar*. PhD thesis, University of Amsterdam.

Van Benthem, J. (1984). Questions about quantifiers. *Journal of Symbolic Logic*, 49(2):443–466.

Van Benthem, J. (1988). Essays in logical semantics. *Studia Logica*, 47(2):172–173.

Van Benthem, J. (1995). *Language in action: categories, lambdas and dynamic logic*. MIT Press.

Van Benthem, J. (2007). Dynamic logic for belief revision. *Journal of Applied Non-Classical Logics*, 17(2):129–155.

Van Benthem, J., Grossi, D., and Liu, F. (2014). Priority structures in deontic logic. *Theoria*, 80(2):116–152.

Van Benthem, J. and Liu, F. (2007). Dynamic logic of preference upgrade. *Journal of Applied Non-Classical Logics*, 17(2):157–182.

Van Benthem, J. and Liu, F. (2018). Deontic logic and changing preferences. *Handbook of Philosophical Logic: Volume 18*, pages 1–49.

Van Benthem, J. and Liu, F. (2020). New logical perspectives on monotonicity. In Deng, D., Liu, F., Liu, M., and Westerståhl, D., editors, *The Second Tsinghua Interdisciplinary Workshop on Logic, Language, and Meaning: Monotonicity in Logic and Language*, volume 12564 of *Lecture Notes in Computer Science*. Springer.

Van Benthem, J., Van Otterloo, S., and Roy, O. (2006). Preference logic, conditionals and solution concepts in games. In Lagerlund, H., Lindström, S., and Sliwinski, R., editors, *Modality Matters: Twenty-Five Essays in Honour of Krister Segerberg*, page 61–77.

Veltman, F. (1996). Defaults in update semantics. *Journal of Philosophical Logic*, 25(3):221–261.

Veltman, F., Groenendijk, J., and Stokhof, M. (1996). This might be it. In Westerståhl, D. and Seligman, J., editors, *Language, Logic, and Computation: the 1994 Moraga Proceedings*, pages 255–70. CSLI.

Villalta, E. (2000). Spanish subjunctive clauses require ordered alternatives. In *Semantics and linguistic theory*, volume 10, pages 239–256.

Von Fintel, K. (1999). NPI licensing, strawson entailment, and context dependency. *Journal of semantics*, 16(2):97–148.

Von Fintel, K. (2012). The best we can (expect to) get? – Challenges to the classic semantics for deontic modals. In *Central Meeting of the American Philosophical Association, February*, volume 17.

Von Fintel, K. (2018). On the monotonicity of desire ascriptions. Manuscript.

Von Fintel, K. and Gillies, A. S. (2008). Cia leaks. *Philosophical review*, 117(1):77–98.

Von Fintel, K. and Gillies, A. S. (2010). Must... stay... strong! *Natural language semantics*, 18:351–383.

Von Wright, G. H. (1967). Deontic logics. *American philosophical quarterly*, 4(2):136–143.

Von Wright, G. H. (1968). *An Essay in Deontic Logic and the General Theory of Action: With a Bibliography of Deontic and Imperative Logic*. Amsterdam: North-Holland Pub. Co.

Westerståhl, D. (2012). Classical vs. modern squares of opposition, and beyond. *The square of opposition. A general framework for cognition*, pages 195–229.

Willer, M. (2013). Dynamics of epistemic modality. *Philosophical Review*, 122(1):45–92.

Willer, M. (2014). Dynamic thoughts on ifs and oughts. *Philosophers' Imprint*, 14:1–30.

Williamson, T. (2000). *Knowledge and its limits*. Oxford University Press.

Wittgenstein, L. (1953). *Philosophical investigations*. Blackwell Publishing.

Xie, K. and Jialiang, Y. (2022). A logic for desire based on causal inference. *Journal of Logic and Computation*. To appear.

Yablo, S. (2010). Permission and (so-called epistemic) possibility. In *Modality*, pages 229–256. Oxford University Press.

Yalcin, S. (2007). Epistemic modals. *Mind*, 116(464):983–1026.

Yan, J., Aloni, M., and Liu, F. (2023). Shifting perspectives: multi-agent extension of BSEL. Manuscript.

Yang, F. and Väänänen, J. (2017). Propositional team logics. *Annals of Pure and Applied Logic*, 168(7):1406–1441.

Zimmermann, T. E. (2000). Free choice disjunction and epistemic possibility. *Natural Language Semantics*, 8:255–290.

Zimmermann, T. E. (2006). Monotonicity in opaque verbs. *Linguistics and Philosophy*, 29(6):715–761.

SAMENVATTING

Modaliteit neemt een centrale positie in binnen de vakgebieden van formele semantiek en modale logica. Dit proefschrift richt zich op epistemische en bouletische modaliteiten en heeft als doel verschillende vraagstukken met betrekking tot monotoniciteit in modale contexten aan te pakken.

In de dissertatie worden twee hoofdonderwerpen onderzocht. Het eerste betreft een systematische en geïntegreerde benadering om puzzels met betrekking tot monotoniciteit in intentionele contexten die zich manifesteren als empirische fenomenen in natuurlijke taal te verklaren, Bij deze puzzels lijken monotone inferenties onder modale en attitudewerkwoorden ongepast te zijn, wat de monotone semantiek van modaliteiten ter discussie stelt. Deze dissertatie suggereert dat deze puzzels voortkomen uit pragmatische effecten die worden veroorzaakt door de typisch ongespecificeerde aard van conclusies die bereikt worden door monotone redenering. In Hoofdstuk 4 wordt een uitgebreide analyse en een uniforme verklaring voor dit probleem voorgesteld.

De tweede focus van dit proefschrift is de studie van modaliteit. Het proefschrift maakt gebruik van epistemische en bouletische modaliteiten als voorbeelden om te onderzoeken of en hoe de puzzels met betrekking tot monotoniciteit invloed hebben op de semantische en logische eigenschappen van de modaliteiten. Ten eerste worden de interacties tussen kennis, overtuigingen en epistemische mogelijkheden (uitgedrukt door het Engelse hulpwerkwoord *might*) systematisch onderzocht. In Hoofdstuk 2 wordt BSEL (Bilateral State-based Epistemic Logic) voorgesteld, waarin de concepten van weten en geloven van een epistemische mogelijkheid worden verkend. Specifiek wordt het fenomeen van epistemische *might* onderzocht en hoe de niet-klassieke eigenschappen ervan interageren met de monotoniciteit van weten en geloven.

In Hoofdstuk 3 breiden we onze discussie uit naar situaties met meerdere subjecten, met als doel te onderzoeken hoe epistemische modaliteiten zich gedragen in multi-agent contexten. Deze uitbreiding wordt gemotiveerd door de waarneming dat epistemische *might* perspectiefgevoelig is. Bijgevolg kunnen verschillende subjecten claims met betrekking tot epistemische might op verschillende manieren evalueren, waardoor de rol van “agentschap” in hun interpretatie moet worden meegenomen.

Hoofdstuk 5 verlegt de focus naar de bouletische modaliteit (wens) en stelt een nieuwe logica van verlangen voor die causale inferentie incorporeert. Deze voorstelling combineert het 'betterness' model van voorkeurslogica met het causale model van causale inferentie, resulterend in een verlangen-causaliteitsmodel. Bovendien is er een volledige logica ontwikkeld voor dit model.

In deze dissertatie wordt de behandeling van monotoniciteit en modaliteit niet los van elkaar uitgevoerd, maar veeleer vanuit het oogpunt van hun interactie. Dit onderzoek biedt inzicht in de relatie tussen logica en taal, semantiek en pragmatiek. De bevindingen dragen bij aan ons begrip van deze onderling verbonden domeinen.

Trefwoorden: monotoniciteit; epistemische mogelijkheid; feitelijkheid van kennis; plausibiliteitsmodellen voor geloof; bilaterale toestandsgebaseerde modale logica; perspectiefgevoeligheid; verlangen en voorkeur; causale inferenties

ABSTRACT

Modality holds a central position in the fields of formal semantics and modal logic. This dissertation delves into epistemic and desiderative modalities, focusing on addressing various monotonicity puzzles in modal contexts.

Two primary topics are investigated throughout the dissertation. The first entails a systematic and unified approach to explain puzzles related to monotonicity in intensional contexts, which manifest as empirical phenomena in natural language. By these puzzles, monotonic inferences under modal and attitude verbs appear infelicitous, thereby challenging the monotonic semantics of modalities. This dissertation suggests that these puzzles arise from pragmatic effects triggered by the typical underspecified nature of conclusions reached by monotonic reasoning. In Chapter 4, a comprehensive analysis and a uniform account for this issue is provided.

Another focus of this dissertation is the study of modality. The thesis employs epistemic and desiderative modalities as examples to examine whether and how the puzzles related to monotonicity influence the semantic and logical properties of the modalities. Firstly, the interactions between knowledge, beliefs, and epistemic possibilities (expressed by the English modal verb *might*) are systematically investigated. In Chapter 2, a Bilateral State-based Epistemic Logic (BSEL) is proposed, exploring the concepts of knowing and believing an epistemic possibility. Specifically, it investigates the phenomenon of epistemic *might* and how its non-classical properties interact with the monotonicity of *know* and *believe*.

In Chapter 3, we broaden our discussion to include situations involving multiple agents, aiming to investigate how epistemic modals interact in multi-agent contexts. This extension is motivated by the argument that epistemic *might* is perspective-sensitive. Consequently, when it comes to claims involving epistemic *might*, different agents may evaluate them differently, thereby requiring the inclusion of agency as a factor in their interpretation.

Chapter 5 shifts its focus towards desiderative modality, proposing a novel logic of desire that incorporates causal inference. This proposal combines the betterness model of preference logic with the causal model of causal inference, resulting in a desire-causality model. Furthermore, a complete logic is developed for this model.

In this dissertation, the treatment of monotonicity and modality is not conducted in isolation, but rather from the standpoint of their interaction. The investigations offer insights into the relationship between logic and language, semantics and pragmatics. These findings contribute to our understanding of these interrelated domains.

Keywords: monotonicity; epistemic *might*; factivity of knowledge; plausibility models for belief; Bilateral State-based Modal Logic; perspective sensitivity; desire and preference; causal inferences

摘 要

模态词一直是形式语义学和模态逻辑领域的重要研究对象。本论文从阐述内涵语境中单调推理的视角出发，对认知和欲望模态词的语义特点及逻辑属性进行了深入研究。

本论文主要探讨两个核心主题。首先，以统一的理论来系统地解释内涵语境中与单调性相关的谜题和悖论。这些谜题似乎阻碍了模态语境中的单调推理，因此对模态词和态度词语义的单调性提出挑战。本论文认为，这一现象源于单调推理所触发的弱化在某些情境下导致了一些语用效应，从而使得单调性推理与自然语言使用之间产生张力，并显示出表面上的非单调性。在第4章中，提出了一种“再解释理论”来全面地解释这个问题。

本论文的另一个研究主题是模态词。研究采用认知和欲望模态和态度动词作为例子，以检验这些单调性谜题是否影响以及如何影响我们对模态词语义的理解和刻画。首先，第2章系统地研究了知识、信念和认知可能性（由英语情态动词 *might* 表达）之间的相互作用。研究还提出了基于状态的双向认知逻辑框架（BSEL），用于探讨知识、信念和认知可能性等概念。具体而言，它研究了与认知可能性的一些经典和非经典性质与知识和信念之间的交互。尤其当认知可能性嵌入在事实性动词“知道”所引领的子句中，知识的事实性便难以被刻画，而 BSEL 框架能很好地解决这些冲突。

在第3章中，我们扩展了模型的讨论范围，使之能够刻画多主体场景。这一扩展的目的是研究认知模态词在多主体环境中的交互。尤其当我们假设认知可能性具有视角敏感性时，不同的主体对同一个包含认知可能性的句子会有不同的估值。因此需要在认知可能性的语义解释中将主体性因素考虑在内。

在第 5 章中，重点转向对欲望模态词的研究，该章节提出将因果推理纳入欲望解释的方法。通过将偏好逻辑的更优模型与因果推理逻辑中的因果模型相结合，构建了一个欲望-因果模型。基于这个模型，提出了一个逻辑系统，并证明了该系统的可靠性和完全性。

这篇论文并没有孤立地研究单调性和模态词，而是以交互的视角作为出发点进行研究。这些研究能够为我们理解逻辑与语言、语义与语用之间的关系提供一个新的视角。

关键词：单调性；认知可能性；知识的事实性；信念的拟真模型；基于状态的双向

模态逻辑；角度敏感性；欲望和偏好；因果推理

RÉSUMÉ AND ACADEMIC ACHIEVEMENTS

Résumé

Jialiang Yan was born on the 22nd of July 1993 in Chengde, Hebei, China.

He began his bachelor's study in the Department of Philosophy, Hebei University in September 2012, majoring in philosophy, and got a Bachelor of Philosophy degree in June 2016.

He began his master's study in the Institute of Philosophy, Chinese Academy of Social Sciences in September 2016, and got a Master of Philosophy degree in Logic in June 2019.

In September 2019, he started to pursue a doctorate in the Department of Philosophy, Tsinghua University. In January 2020, he was admitted to the jointly awarded doctorate program of Tsinghua University and the Institute for Logic, Language and Computation, University of Amsterdam.

Academic Achievements

[1] Kaibo Xie, Jialiang Yan. A logic for desire based on causal inference. Journal of Logic and Computation. Accepted (extended journal version of the LORI21 conference paper). 2022.

[2] Jialiang Yan. The overtone of monotonicity under desire. Proceedings of the ESSLLI 2022 Student Session. 2022, Published online: https://2022.esslli.eu/courses-workshops-accepted/student-session.

[3] Jialiang Yan, Fenrong Liu. Monotonic opaqueness in deontic contexts. In Proceedings of 5th AWPL, Liao B., Wáng Y. (eds) Context, Conflict and Reasoning. Logic in Asia: Studia Logica Library. Springer, Singapore. pp: 87-96. 2022.

[4] Jialiang Yan. Review of *Nominals of Possible Worlds*. Studies in Logic, 2020(5): 97-100. 2020 (published in Chinese).

[5] Jialiang Yan, Maria Aloni. Knowing and believing an epistemic possibility. *Synthese*. Under review. 2023.

[6] Jialiang Yan, Maria Aloni and Fenrong Liu. Shifting perspectives: an multi-agent extension of BSEL. Manuscript. 2023.

Titles in the ILLC Dissertation Series:

ILLC DS-2018-04: **Jelle Bruineberg**
Anticipating Affordances: Intentionality in self-organizing brain-body-environment systems

ILLC DS-2018-05: **Joachim Daiber**
Typologically Robust Statistical Machine Translation: Understanding and Exploiting Differences and Similarities Between Languages in Machine Translation

ILLC DS-2018-06: **Thomas Brochhagen**
Signaling under Uncertainty

ILLC DS-2018-07: **Julian Schlöder**
Assertion and Rejection

ILLC DS-2018-08: **Srinivasan Arunachalam**
Quantum Algorithms and Learning Theory

ILLC DS-2018-09: **Hugo de Holanda Cunha Nobrega**
Games for functions: Baire classes, Weihrauch degrees, transfinite computations, and ranks

ILLC DS-2018-10: **Chenwei Shi**
Reason to Believe

ILLC DS-2018-11: **Malvin Gattinger**
New Directions in Model Checking Dynamic Epistemic Logic

ILLC DS-2018-12: **Julia Ilin**
Filtration Revisited: Lattices of Stable Non-Classical Logics

ILLC DS-2018-13: **Jeroen Zuiddam**
Algebraic complexity, asymptotic spectra and entanglement polytopes

ILLC DS-2019-01: **Carlos Vaquero**
What Makes A Performer Unique? Idiosyncrasies and commonalities in expressive music performance

ILLC DS-2019-02: **Jort Bergfeld**
Quantum logics for expressing and proving the correctness of quantum programs

ILLC DS-2019-03: **András Gilyén**
Quantum Singular Value Transformation & Its Algorithmic Applications

ILLC DS-2019-04: **Lorenzo Galeotti**
The theory of the generalised real numbers and other topics in logic

ILLC DS-2019-05: **Nadine Theiler**
Taking a unified perspective: Resolutions and highlighting in the semantics of attitudes and particles

ILLC DS-2019-06: **Peter T.S. van der Gulik**
Considerations in Evolutionary Biochemistry

ILLC DS-2019-07: **Frederik Möllerström Lauridsen**
Cuts and Completions: Algebraic aspects of structural proof theory

ILLC DS-2020-01: **Mostafa Dehghani**
Learning with Imperfect Supervision for Language Understanding

ILLC DS-2020-02: **Koen Groenland**
Quantum protocols for few-qubit devices

ILLC DS-2020-03: **Jouke Witteveen**
Parameterized Analysis of Complexity

ILLC DS-2020-04: **Joran van Apeldoorn**
A Quantum View on Convex Optimization

ILLC DS-2020-05: **Tom Bannink**
Quantum and stochastic processes

ILLC DS-2020-06: **Dieuwke Hupkes**
Hierarchy and interpretability in neural models of language processing

ILLC DS-2020-07: **Ana Lucia Vargas Sandoval**
On the Path to the Truth: Logical & Computational Aspects of Learning

ILLC DS-2020-08: **Philip Schulz**
Latent Variable Models for Machine Translation and How to Learn Them

ILLC DS-2020-09: **Jasmijn Bastings**
A Tale of Two Sequences: Interpretable and Linguistically-Informed Deep Learning for Natural Language Processing

ILLC DS-2020-10: **Arnold Kochari**
Perceiving and communicating magnitudes: Behavioral and electrophysiological studies

ILLC DS-2020-11: **Marco Del Tredici**
Linguistic Variation in Online Communities: A Computational Perspective

ILLC DS-2020-12: **Bastiaan van der Weij**
Experienced listeners: Modeling the influence of long-term musical exposure on rhythm perception

ILLC DS-2020-13: **Thom van Gessel**
Questions in Context

ILLC DS-2020-14: **Gianluca Grilletti**
Questions & Quantification: A study of first order inquisitive logic

ILLC DS-2020-15: **Tom Schoonen**
Tales of Similarity and Imagination. A modest epistemology of possibility

ILLC DS-2020-16: **Ilaria Canavotto**
Where Responsibility Takes You: Logics of Agency, Counterfactuals and Norms

ILLC DS-2020-17: **Francesca Zaffora Blando**
Patterns and Probabilities: A Study in Algorithmic Randomness and Computable Learning

ILLC DS-2021-01: **Yfke Dulek**
Delegated and Distributed Quantum Computation

ILLC DS-2021-02: **Elbert J. Booij**
The Things Before Us: On What it Is to Be an Object

ILLC DS-2021-03: **Seyyed Hadi Hashemi**
Modeling Users Interacting with Smart Devices

ILLC DS-2021-04: **Sophie Arnoult**
Adjunction in Hierarchical Phrase-Based Translation

ILLC DS-2021-05: **Cian Guilfoyle Chartier**
A Pragmatic Defense of Logical Pluralism

ILLC DS-2021-06: **Zoi Terzopoulou**
Collective Decisions with Incomplete Individual Opinions

ILLC DS-2021-07: **Anthia Solaki**
Logical Models for Bounded Reasoners

ILLC DS-2021-08: **Michael Sejr Schlichtkrull**
Incorporating Structure into Neural Models for Language Processing

ILLC DS-2021-09: **Taichi Uemura**
Abstract and Concrete Type Theories

ILLC DS-2021-10: **Levin Hornischer**
Dynamical Systems via Domains: Toward a Unified Foundation of Symbolic and Non-symbolic Computation

ILLC DS-2021-11: **Sirin Botan**
Strategyproof Social Choice for Restricted Domains

ILLC DS-2021-12: **Michael Cohen**
Dynamic Introspection

ILLC DS-2021-13: **Dazhu Li**
Formal Threads in the Social Fabric: Studies in the Logical Dynamics of Multi-Agent Interaction

ILLC DS-2022-01: **Anna Bellomo**
Sums, Numbers and Infinity: Collections in Bolzano's Mathematics and Philosophy

ILLC DS-2022-02: **Jan Czajkowski**
Post-Quantum Security of Hash Functions

ILLC DS-2022-03: **Sonia Ramotowska**
Quantifying quantifier representations: Experimental studies, computational modeling, and individual differences

ILLC DS-2022-04: **Ruben Brokkelkamp**
How Close Does It Get?: From Near-Optimal Network Algorithms to Suboptimal Equilibrium Outcomes

ILLC DS-2022-05: **Lwenn Bussière-Carae**
No means No! Speech Acts in Conflict

ILLC DS-2023-01: **Subhasree Patro**
Quantum Fine-Grained Complexity

ILLC DS-2023-02: **Arjan Cornelissen**
Quantum multivariate estimation and span program algorithms

ILLC DS-2023-03: **Robert Paßmann**
Logical Structure of Constructive Set Theories

ILLC DS-2023-04: **Samira Abnar**
Inductive Biases for Learning Natural Language

ILLC DS-2023-05: **Dean McHugh**
Causation and Modality: Models and Meanings