Pharaoh: Conceptual Blending of Cognitive Scripts for Computationally Creative Agents

[1,2]Rania Hodhod, [1]Brian Magerko

[1]ADAM Lab, Georgia Institute of Technology, 85 5th Street, Atlanta, Georgia 30332 USA
[2]Faculty of Computer and Information Sciences, Ain Shams University, Abbassia, Cairo 11566 Egypt
{rhodhod6; magerko}@gatech.edu

Abstract

Improvisational acting is a creative group performance where actors co-construct stories on stage in real-time based on actors' perceptions of the environment. The Digital Improv Project has been engaged in a multi-year study of the cognitive processes involved in improvisational acting. This better understanding of human cognition and creativity has led to formal computational models of some aspects of our findings. In this work, we consider enriching AI improv agents with the ability to improvise new nontraditional scenes based on existing social cognitive scripts. This paper shows how the use of Pharaoh -a context based structural retrieval algorithm for cognitive scripts- and simple blending rules can help digital improv agents to create new interesting scenes. The paper also provides an illustrative example at the end.

Introduction

Improvisational acting is a creative group performance where actors co-construct stories on stage in real-time based on actors' perceptions of the environment. We have seen in our socio-cognitive studies of improvisational actors that human improvisers often use their knowledge and past experience to co-create a new scene and that cognitive scripts are often employed in the co-creation of improvised scenes (Magerko et al., 2009; Magerko & O'Neill, 2012). Cognitive scripts are the symbolic basis for representing behavioral task and domain knowledge. They are defined as a recurrent spatial structure that gives coherence to our experience (Johnson, 1987). These scripts are not replayed verbatim, however, and actors often use other processes to vary, modify, and combine scripts to create new ideas and experiences for an improvised scene. The question, in this conception, becomes: how can we enable AI agents to create new interesting improvised scenes?

Virtual improvisational theatre has been the focus of research in digital improv systems. The Computer-Animated Improvisational Theater (CAIT) offered two types of improv agents: Improv Puppets and Improv Actors (Li, Lee-Urban, Appling, & Riedl, 2012; Nelson & Mateas, 2005). Both types of agents could improvise with other agents in real-time under the guidance of human input for how they should act. These agents reasoned about each other's mental models (i.e. an explanation of someone's thought process about how something works in the real world) and could change their own mental model to conform to another's based on observations. However, CAIT cannot allow humans to improvise with these systems as equal co-creators of a narrative because these agents could not create narratives on their own as a result of the absence of the procedural knowledge required for this act. Instead, they worked from pre-authored narrative structures. In the Improv system (Sharma, Ontanon, Mehta, & Ram, 2010), virtual animated avatars could be scripted to enact a scenario giving the appearance of believable improvisational performances. However, none of these systems applied a deep understanding of how professional actors perform improvisation, nor have any of these systems provided the improv agents with the ability to 'independently' create new scenes, which is the focus of this paper.

As we mentioned earlier, improvisational acting requires past experience and successful retrieval of related experiences to be reused or blended in a way that enriches the current experience. This paper describes the use of cognitive scripts as social cognitive constructs and Pharaoh (a context-based structural retrieval algorithm) (Hodhod, Magerko, & Gawish, 2013) to allow improv agents to recall past similar experiences and utilize them to construct new scenes.

Cognitive Scripts as Social Cognitive Constructs

Consider the following story adopted from Ritchie (2004) which can be reasonable example for an improviser's narrative experience: "He used what he thought was a fake gun in the holdup but it turned out that it really was a gun and the clerk behind the counter was an under-cover police officer so he was charged with armed robbery and assaulting a police officer." Despite the fact that this story is short, it can be interpreted as a complex narrative and would require the creation of a long sequence of entirely separate mental spaces (representations) (Ritchie, 2004). However, these independent mental spaces are dissolved as new ones are generated, so that the load on cognitive capacity would not expand rapidly.

One way to conceptualize the above story is in the form of cognitive scripts. A cognitive script is a schema that describes an organized pattern of thought or behavior. Accordingly, the first script in the above story represents the common robbery script, then altered as the narrative progresses by adding or changing connections with other scripts (e.g., for fake gun and under-cover policeman) and finally with a criminal trial script. We found that such stories provide a good source of common social experiences that can be used to build our corpus of scripts. Accordingly, we collected a number of stories in the form of cognitive scripts using Mechanical Turk and saved them in one corpus. These scripts aim to provide the agent with various social experiences.

Moreover, the above story also reflects an interesting blend in which the real bank robbery script is blended with other scripts where 'fake gun' was used instead of 'real gun' and 'under-cover police officer' was used instead of 'real clerk', which is exactly what we would like our improv agent to be able to do, that is combining several scripts into a new meaningful, hopefully intersting, script.

Although cognitive scripts reveal very little about how the actions of a sequence are causally chained together and do not display the roles of objects in temporal sequences (Chen, 2004), they successfully describe a stereotyped sequence of events in a particular context (events can be linear or multi-branched) (Luger & Stubblefield, 1998). In fact, it has been shown that this temporal sequence of events provided in cognitive scripts helps in defining the context of those scripts (Hodhod, Magerko, & Gawish, 2013), which can be accordingly used by improv agents to define those scripts that share similar context with each other and consequently can make use of those scripts to come up with a new blended script, similar to the kind of blend occurred in the above story.

Improvisational Acting

In an improvisational act, each improviser starts to act and build an individual mental model based on how he perceives the world around him (Magerko et al., 2009) and is guided only by sets of game rules about how a scene should be functionally performed (Johnstone, 1999; Spolin, 1963). Gradually, along the scene, improv actors start to develop common beliefs forming what is known as a shared mental model. Shared mental models (SMMs) are cognitive constructs that contain common beliefs of the team members (Hodhod & Magerko, 2012) maintained through the process of reaching cognitive convergence; that is two or more team members become on the same page.

Improv actors normally negotiate their shared mental model (i.e. shared common beliefs about the improvised world) solely through performative actions in the ongoing scene rather than with explicit communication about the model (Fuller & Magerko, 2010). When improvisers recognize divergences they aim to repair them (i.e. attempting to converge on a shared mental model), and accept the repair (i.e. seeing the result of the attempted repair) (Fuller & Magerko, 2010).

The skills necessary for expert improvisational acting are the result of years of training and performance plus a lifetime of real world experience. Improv actors use their past experiences and knowledge to construct interesting blended stories. These stories can be easily described in the form of partial ordering of events that defines a space of possible event sequences that can unfold during a given situation called scripts. Improvisers rely on their experience to pick an appropriate script from their long term memory to improvise. Generally, this may involve swinging back-and-forth between varieties of social cognitive scripts, hence enriches the actor's current experience.

Improvisers use blending to create new interesting scenes by using one or more social cognitive scripts. Since it is easy for human improvisers to recognize the resemblance between the dating script and interview script, they can borrow/adapt events/structures from one script and use them in another script. For example, a human improviser can easily create an interesting situation when acting the role of a boyfriend in a dating script by asking his girlfriend if she knows how to use a typewriter. Furthermore, an improviser who is acting the role of an interviewer in an office script might ask the interviewee for the reason she left her ex-boyfriend. These unusual events might lead to more interesting events while the scene unfolds. Such blend can be easily achieved by humans, but is not currently achievable by digital improv agents.

In this paper, we present a two steps technique that allows digital improv agents to exhibit creative behavior. The first step involves the use of Pharaoh to allow an agent to retrieve similar contextually structured cognitive scripts from a corpus of scripts. The second step involves the application of blending rules that allow the agent to borrow and use events/structures from the retrieved cognitive script as part of their current experience.

Pharaoh (Context-Based structural retrieval Algorithm)

This work aims to use Pharaoh and cognitive social scripts to foster creativity in improvisational agents. Our aim is to allow the improv agents to make use of past experiences and come up with new experiences. Blending is one way to obtain new and/or interesting experiences. Blending requires the presence of two input domains, in our case, the current experience (Target) and a retrieved past experience (Source). Isomorphism (one-to-one mapping) is an essential condition in the blending process as identified by Fauconnier and Turner (1996), therefore we hypothesized that the best script would be the one that holds corresponding events to the current script. However, isomorphism is not enough by itself to compute similarity between Target and Source. Context is another important factor that needs to be considered in the retrieval process. For example, a ‘fight' event that occurred in the beginning of a script might not hold the same context of a ‘fight' event that occurred in the end of another script. Therefore, it is vital to consider the temporal occurrence of events in both scripts along in the computation of the similarity measure.

Pharaoh is a context-based structural retrieval algorithm that allows retrieval of similar cognitive scripts to a query script. Pharaoh uses cognitive scripts in which each node represents an event associated with all of the objects involved in that event. The structure of the script can be understood by examining the components of those “nodes.” A node consists of a ‘node id', ‘relation name', ‘parameters', and ‘children nodes'. A script composed of a set of *before* relations, B(e_1, e_2), between events e_1 and e_2 signifying that e_1 occurs before e_2, in the sense that “what is in one slot affects what can be in another” (Chen, 2004) (see Figure 1 for script examples). These *before* relations coincide with causal and temporal precedence information, which are important for narrative comprehension (Graesser, Singer, & Trabasso, 1994). Pharaoh uses the temporal links to compute the similarity between individual nodes as part of the final similarity between the two cognitive scripts under consideration.

Since pharaoh will be used by the improv agent in real time, we dedicated the next section to explain the run time complexity of Pharaoh.

Pharaoh: Run-Time Complexity

The complexity of Pharaoh is O(n^2) in the best case, and O(n^4) in the worst case. The computational complexity of Pharaoh is the result of computing the similarity between two cognitive scripts and depends on three factors: (i) the number of actions in each of the plot graphs, (ii) the time needed to compute similarity between corresponding nodes (actions), and (iii) the time needed to find and filter corresponding attributes. The following paragraph gives more detail about these three factors.

For the first factor, we can assume that the plot graphs for Target and Source consist of *n* nodes (events) each. Therefore, the cost of the first factor will be O(n^2); since we compare each node in Target to each node in Source. The cost of the second factor can be decomposed into two terms: a) the cost needed to check if nodes share the same relations which takes constant time *c* and the cost needed to find the least common parent (LCP) of the corresponding nodes which is accomplished by using breadth first search (BFS-up) –from the nodes up to their LCP– this search is done with a cost proportional to the depth of the plot graph, that is O($|E|+|V|$) (Cormen, Leiserson, Rivest & Stein, 2001). b) the total cost of the second factor will be ($c + 2\ |E| + 2\ |V|$) with complexity O(n), where $|E| = n\text{-}1$ and $|V| = n$, considering the validity of the assumption of the first factor. Finally, the cost of the third factor is the sum of the cost of finding the corresponding attributes m, i.e. O(m). Consequently, the total cost of the third factor will be O(m) where m is small.

Illustrative Example

This section explains how Pharaoh provides an agent with a context similar script (Source) that it can use/blend with the current script (Target) to produce an interesting scene/act. Pharaoh considers Target as the query script and starts to look for all similar scripts from the set of all existing scripts. Pharaoh computes similarities between Target and all other scripts using context, lexically-similar words and similar structure between Target and each of the scripts in the agent's memory. The most similar script (i.e. script with highest similarity) is retrieved as Source, see Table 1. For more information on how Pharaoh works please see (Hodhod, Magerko, & Gawish, 2013).

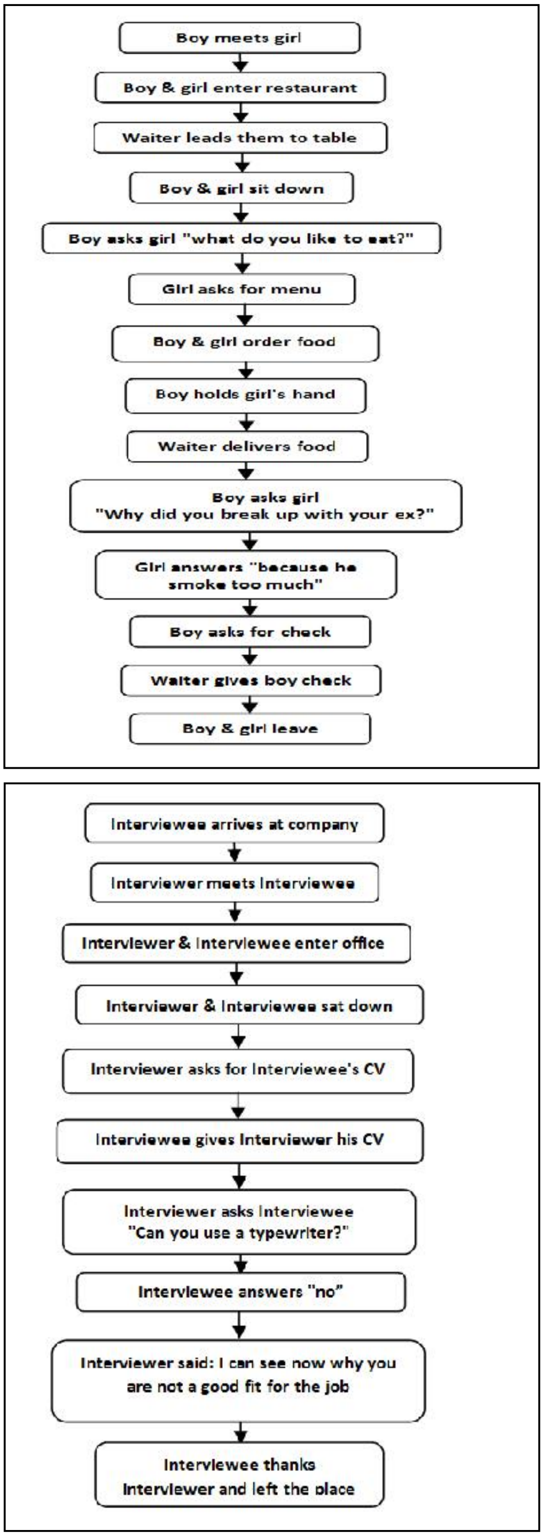

Figure 1. Linear paths from the dating and interview scripts

Table1. Similarity measures between dating script and all other existing scripts.

Script	Cinema	Interview	Concert	Buying a Car	Stadium
Dating	0.11	1.77	0	0	0.13

Table 1 shows the similarity measures between Target (dating script) and all other existing scripts in the agent's memory. Pharaoh retrieved *interview* as the most similar script since it has the highest similarity value. Target and Source can be used now by the improv agent to construct a blended scene using projections from both scripts. According to Fauconnier and Turner (1998) projections to the blended space are partially selective; not all elements from the inputs are projected to the blend. This requires the presence of projection rules that decide on what projections stay in the blended space and what would be disregarded. In the meantime, we are using a rule that was used before in the cognitive system, Sapper (Veale & O Donoghue, 2000) , that is to keep the structure (skeleton) of one script and either move new events to that script or change some constructions of other events, such as replacing drinks with tickets in the buying event. The effect of applying this rule can be seen in Figure 2.

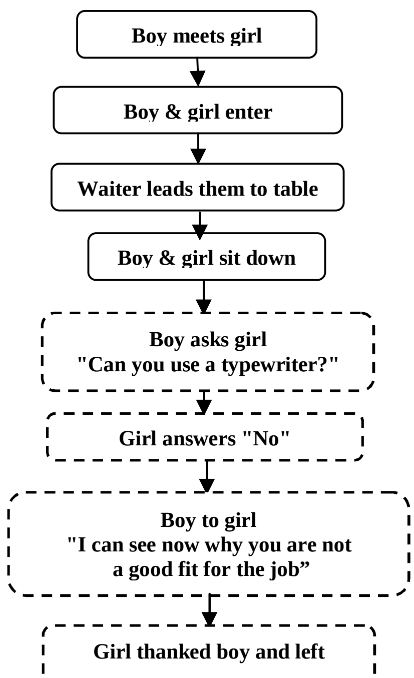

Figure2. Blended script

The blend in Figure 2 tells the following story: couple goes to the restaurant; boyfriend asks girlfriend "what do you like to eat"; girlfriend asks for the menu; boyfriend holds girlfriend's hand; boyfriend asks girlfriend "can you use a typewriter?", girlfriend: "no", boyfriend: "I can see now why you are not a good fit for the job", girlfriend thanks boyfriend and left.

What happened in the new script is that the agent decided on when to start the blend by looking for a least common parent node between Target and Source (in this case, the 'sat-down' event). All the events below this node are transferred to Target, but with the substitution of the mapped attributes in those events. For example, 'interviewee' is substituted by 'girl' and 'interviewer' is substituted by 'boy'. Further blending can take place by setting the new blend as Target and look for another Source that can be used to extend the script, if needed.

Discussion

Imagine an improvised scene in which two improv actors adopt two different scripts and each actor is following his script (i.e. they are not on the same page yet). In order for the actors to achieve consensus, a conceptual change is required in which one actor needs to agree with the other actor via adopting his script. Such conceptual change depends on two factors: first, the extent to which the two relevant scripts differentiate from each other; and second, what kinds of cognitive mechanisms are available for eliminating these differences (Chen, 2004). Pharaoh can successfully handle the first factor. Pharaoh is able to provide the similarity values for each script in the corpus and the query script, in addition to mapping corresponding events. The second factor proposed by Chen is addressed via the use of repair strategies that allow the actors to get rid of the current divergence and reach cognitive consensus (Fuller & Magerko, 2010; Hodhod & Magerko, 2012; Magerko, Dohogne, & Fuller, 2011). Allowing improv agents to possess those capabilities would be considered a big step not only in field of interactive narrative, but more broadly in computational creativity.

To the extent of our knowledge, retrieval of cognitive scripts from a corpus of scripts provides a challenge in the information retrieval area in general and analogical retrieval in particular (Hodhod, Magerko, & Gawish, 2013). The challenge appears in the temporal precedence information held in cognitive scripts and which should be considered in the retrieval process in order to preserve the social context of the scripts. Retrieval of cognitive scripts is the first step in achieving a computational blending model for social constructs.

Studies and analyses of blending have been carried out in areas such as sign language (Liddell, 2000), music theory (Zbikowski, 1999), cinema (Veale, 1996), humor (Coulson, Urbach, & Kutas, 2006), design (Ribeiro et al., 2003), and poetry (Harrell, 2007). However, blending studies is underestimated in the improvisation domain. This paper shows how Pharaoh can stretch the capabilities of digital improv agents and highlights the contributions of this work over existing ones. Our findings can be used to gain a more comprehensive understanding of the practical construction of cognitive social scripts and the blending process might occur on them as well as draw attention to techniques worthy of further examination.

In the study conducted to compare the performance of Pharaoh to that of humans, Pharaoh showed 75% agreement with the human participants' selections. We are currently investigating the possibility of enhancing Pharaoh's performance using heuristics. Pharaoh doesn't seem to work efficiently with large size scripts (large number of branches) as it takes Pharaoh long time to compute the similarity values. On the other hand, with reasonable sized scripts, Pharaoh can be considered an effective module that provides an AI improv agent with the second mental space required for the blending process.

Conclusion

Digital improvisation is a complex and challenging domain where many mental processes are involved as such understanding and negotiation of mental models, recalling of past experiences and blending two or more of them to produce/use in a new situation. Enriching digital improv agents with mental capabilities, such as blending has not been investigated before. This article presents Pharaoh that allows retrieval of cognitive scripts from existing corpus of cognitive scripts. Pharaoh includes a structural mapping mechanism that allows analogical mappings between objects in different domains.

Pharoah, provides a mean to find the most similar cognitive script (Source) to a query one (Target), which is actually an essential step required for blending of cognitive scripts. The proposed technique including Pharaoh and the simple blending rules can act as a primary cognitive model that fosters creativity at digital improv agents. We believe that this work should have an impact not only on digital improvisation and AI-based narrative systems but also on cognitive systems in general. Future work includes investigating the completion and elaboration phases that can follow any blending process (Fauconnier & Turner, 1998) through applying more blending or using common sense knowledge bases as sources for new events.

References

Chen, X. (2004). Scripts and Conceptual Change. In P. Li, X. Chen, and H.X. Zhang (Eds.). *In Science, Cognitive, and Consciousness*, Nanchang, 96-117.

Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2001). *Introduction to Algorithms*. MIT press, pp. 531-534.

Coulson, S., Urbach, T. P., & Kutas, M. (2006). Looking back: Joke comprehension and the space structuring model. *Humor–International Journal of Humor Research*, *19*(3), 229–250.

Fauconnier, G., & Turner, M. (1996). Blending as a central process of grammar. In Goldberg, A. E. (Ed.). *Conceptual structure, discourse, and language*. Stanford, CA: CSLI Publications, 113-131.

Fauconnier, G., & Turner, M. (1998). Conceptual integration networks. *Cognitive science*, *22*(2), 133–187.

Fuller, D., & Magerko, B. (2010). Shared mental models in improvisational performance. *Proceedings of the Intelligent Narrative Technologies III Workshop*, INT3 '10 (pp. 1–6). Monterey, CA: ACM.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review; Psychological Review*, *101*(3), 371.

Harrell, D. F. (2007). GRIOT's Tales of Haints and Seraphs: A Computational Narrative Generation System. In N. Wardrip-Fruin and P. Harrigan (Eds.). *Second Person: Role-Playing and Story in Games and Playable Media*. Cambridge, MIT Press.

Hodhod, R., & Magerko, B. (2012). Reaching Cognitive Consensus with Improvisational Agents. *Proceedings of the 8th Annual AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment* (pp. 6). Palo Alto, CA.

Hodhod, R., Magerko, B., & Gawish, M. (2013). Pharaoh: Context-Based Structural Retrieval of Cognitive Scripts. *International Journal of Information Retrieval Research (IJIRR)*, *2*(3), 58–71.

Hodhod, R., Piplica, A., & Magerko, B. (2012). A formal architecture of shared mental models for computational improvisational agents. *Intelligent Virtual Agents* (pp. 440–446).

Johnson, M. (1987). The body in the mind: The bodily basis of imagination, reason, and meaning. *The body in the mind: the bodily basis of imagination, reason and meaning*. Chicago: University of Chicago Press.

Johnstone, K. (1999). *Impro for Storytellers* (1st ed.). New York: Routledge.

Liddell, S. K. (2000). 16 Blended spaces and deixis in sign language discourse. *Language and gesture*, *2*, Cambridge: Cambridge University Press, 331 - 357.

Luger, G. F., & Stubblefield, W. A. (1998). *Artificial Intelligence: Structures and Strategies for Complex Problem Solving*. Addison-Wesley.

Magerko, B., Dohogne, P., & Fuller, D. (2011). Shared Mental Models in Improvisational Digital Characters. *Proceedings of the 2nd International Conference on Computational Creativity* (pp. 33–35). Presented at the ICCC 2011, Mexico City, Mexico.

Magerko, B., Manzoul, W., Riedl, M., Baumer, A., Fuller, D., Luther, K., & Pearce, C. (2009). An Empirical Study of Cognition and Theatrical Improvisation. *Proceeding of the Seventh ACM Conference on Creativity and Cognition* (pp. 117–126). New York: ACM.

Magerko, B., & O'Neill, B. (2012). Formal Models of Western Films for Interactive Narrative Technologies. *the LREC 2012 Workshop on Computational Models of Narrative*. Istanbul, Turkey.

Nelson, M. J., & Mateas, M. (2005). Search-Based Drama Management in the Interactive Fiction Anchorhead. *Proceedings of the First Annual Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE-05)* (pp. 99–104).

Ribeiro, P., Pereira, F. C., Marques, B., Leitao, B., Cardoso, A., Polo, I. I., & de Marrocos, P. (2003). A model for creativity in creature generation. *Proceedings of the 4th Conference on Games Development (GAME ON'03)* (pp. 170–175).

Ritchie, L. D. (2004). Lost in" Conceptual Space": Metaphors of Conceptual Integration. *Metaphor and Symbol*, *19*(1), 31–50.

Sharma, M., ONTAnon, S., Mehta, M., & Ram, A. (2010). Drama management and player modeling for interactive fiction games. *Computational Intelligence*, *26*(2), 183–211.

Spolin, V. (1963). *Improvisation for the theater: A handbook of teaching and directing techniques*. Northwestern Univ Pr.

Veale, T., & O'Donoghue, D. (2000). Computation and Blending. *Cognitive Linguistics 11*, 3(4) 253–282.

Veale, T. (1996). Pastiche: A Metaphor-centred Computational Model of Conceptual Blending, with special reference to Cinematic Borrowing. *Manuscript.[Available from http://blending. stanford. edu.]*.

Zbikowski, L. M. (1999). The blossoms of'Trockne Blumen': Music and text in the early nineteenth century. *Music analysis*, *18*(3), 307–345.