The Knowledge-Fitting Theory of Plausibility

Louise Connell and Mark T. Keane [1]

Abstract. Plausibility is judged on a daily basis in a wide range of cognitive phenomena, yet the study of plausibility in its own right has been long neglected in cognitive science. In this paper, we present the Knowledge-Fitting Theory of Plausibility that incorporates both concept-coherence (i.e. the conceptual structure and relatedness of a scenario) and word-coherence (i.e. the distributional properties of the individual words used to describe a scenario) in plausibility judgement. We also present the Plausibility Analysis Model (PAM), which is an implementation of this theory and the first computational mode to specifically address the issue of human plausibility judgements.

1 INTRODUCTION

Plausibility is an ineluctable phenomenon of everyday life – engaged in everything from assessing the quality of a movie plot to considering a child's excuse for a broken dish. It is perhaps this very ubiquity that has led to its neglect in cognitive science. Typically, in cognitive science literature, plausibility is merely operationalised (as ratings on a scale), rather than explained. This literature has shown the cognitive processes that utilise plausibility are many and diverse. People often use plausibility judgements in place of costly retrieval from long-term memory, especially when verbatim memory has faded [1][2][3][4]. Plausibility is also used as a kind of cognitive shortcut in reading, to speed parsing and resolve ambiguities [5][6][7]. In everyday thinking, plausible reasoning that uses prior knowledge appears to be commonplace [8], and can even aid people in making inductive inferences about familiar topics [9]. It has also been argued that plausibility plays a fundamental role in understanding novel word combinations by helping to constrain the interpretations produced [10][11]. In this way, the empirical literature leaves us with a sense that plausibility is important but without a good indication of what is actually involved. From this overview it is clear that plausibility is in need of a thorough theoretical, computational and empirical treatment.

In the rest of this paper, we outline the evidence for the effect of two factors on plausibility – concept-coherence and word-coherence. We then detail our Knowledge-Fitting Theory of Plausibility, which proposes an innovative approach to plausibility that incorporates both concept-coherence (i.e. the conceptual relatedness of a described scenario with prior knowledge) and word-coherence (i.e. the distributional information of the individual words used to describe the scenario). In addition, we describe the implementation of the theory in the form of PAM, the Plausibility Analysis Model, and discuss its performance in modelling human data [12][13].

[1] Department of Computer Science, University College Dublin, Belfield, Dublin 4, Ireland. Email louise.connell@ucd.ie

2 EVIDENCE OF EFFECTS

2.1 Concept-Coherence

Although few researchers have expounded a theory of plausibility, there is a shared view that plausibility has something to do with *concept-coherence* – i.e. that "something is plausible if it is conceptually supported by prior knowledge" [8]. For example, it has been demonstrated the importance of concept-coherence to perceived plausibility by disrupting the causal sequence of events in short stories [14]. People's ability to recall these stories, and the plausibility ratings they gave, was sensitive to the degree to which the overall concept-coherence of the story had been altered. Indeed, these re-ordered stories often appeared to be just incomprehensible.

A concept-coherence view of plausibility suggests that when people make a plausibility judgement they relate a target description to their prior experience, and in some way assess whether the current scenario fits in with what they have experienced in the past. If a person read the statement "*The bottle rolled off the shelf and smashed on the floor*" they might make the bridging inference that the bottle falling *caused* it to smash on the floor. This might lead them to judge this scenario as being highly plausible because prior experience tells them that fragile things often break when they fall on hard surfaces. Put simply, the description has a certain conceptual coherence. In contrast, if the target description was "*The bottle rolled off the shelf and melted on the floor*", the person might consider it less plausible because their past experience has few examples of falling fragile objects melting on contact with floors – though a scenario could be construed where this could occur, such as if the room was made of metal and heated up like an oven). In other words, this description lacks a certain conceptual coherence.

It has been shown that manipulating the concept-coherence of a scenario (i.e. by inviting different bridging inferences) affects its perceived plausibility [12][15][16]. For example, sentence pairs linked by causal inferences (*causal pairs* such as the bottle-smashing scenario) were judged as being more plausible than sentences that fail to invite obvious causal inferences (*unrelated pairs* such as the bottle-melting scenario). Furthermore, causal pairs were also found to be more plausible than sentence pairs that invited simple attributal inferences (i.e. where some attribute of the object is mentioned in the second sentence, such as "*The bottle was pretty*"), which in turn were judged to be more plausible than inferences of temporal succession (e.g. "*The bottle sparkled*"). It has also shown that concept-coherence affects the time it takes people to make a binary (yes / no) plausibility judgement [13]. People took significantly longer to make a yes / no decision of plausibility for causal sentence pairs than attributal sentence pairs. These studies provide specific concrete evidence that plausibility is influenced by the conceptual coherence of a situation, as shaped by the type of inferences involved.

2.2 Word-Coherence

While concept-coherence has been seen as an overarching view of what is involved in many aspects of cognition, including plausibility (e.g. [17]), very little consideration has been given to other factors that may influence plausibility judgements. More recently, evidence has been provided for the role of word-coherence in determining plausibility [18]. This view suggests that plausibility judgements are sensitive to the distributional information of the individual words used to describe a scenario. In other words, the distinctive relationship between words, as represented in distributional knowledge, can make certain scenarios seem more plausible simply by virtue of the particular words used.

Distributional knowledge of a language is gleaned through statistical analysis of large corpora that determine how each word in the language is used in relation to every other word. By moving through the corpus and counting the frequency with which a given word appears with other words in its surrounding context, a picture of the distribution of a language is formed. In this fashion, a word can be summarised as a vector – or point in high-dimensional space – showing the frequency with which is it associated with other lexemes in the corpus. Similarly, a sentence may be represented as a single point in distributional space by merging the points of individual words; for example, the Latent Semantic Analysis model (LSA; [19]) uses the weighted sum of constituent word vectors to denote tracts of text. In this way, two sentences containing words that occur in similar linguistic contexts (i.e., that are distributionally similar) will be positioned closer together in this space than two sentences containing words that do not share as much distributional information.

Manipulating the word-coherence of a description has been shown to affect the time needed to decide if a situation is plausible [13]. For example, consider these sentence pairs:

(i) *The pack saw the fox. The hounds snarled.*
(ii) *The pack saw the fox. The hounds growled.*

While (i) and (ii) essentially have the same meaning (i.e. they both invite the same inference), the different distributional properties of *snarled* and *growled* mean that the sentences of pair (i) are further apart distributionally than the sentences of pair (ii) (see Figure 1). People are faster to judge as plausible sentences of type (i) that are distributionally far apart, than sentences of type (ii) that are distributionally close together. So, word-coherence has an effect on plausibility, albeit weaker than that of concept-coherence.

3 THE KNOWLEDGE-FITTING THEORY OF PLAUSIBILITY

3.1 Making a plausibility judgement

In essence, we see plausibility as being about making what one is told fit what one knows about the world. When we ask people to make a plausibility judgement, we are essentially asking them to estimate the goodness of this fit. It is this process that we capture in our Knowledge-Fitting theory of plausibility, which is characterised by two main stages: the comprehension and assessment stages. The *comprehension* stage constructs a representation of a scenario with reference to distributional and prior knowledge. The *assessment* stage analyses this representation to ascertain its fit to prior knowledge, hence assessing its plausibility (see Figure 2). We shall now look at each of these stages in turn.

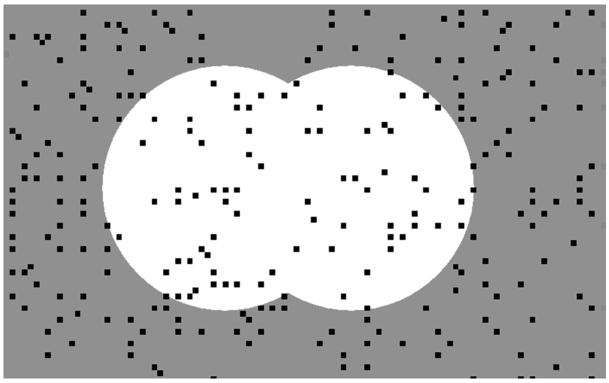

Figure 1. Illustration of the distributional spotlight; sentences that are close togher have overlapping spotlights and give smaller coverage than sentences that are far apart

3.1.1 Comprehension

Our view of the comprehension stage is not much different to that currently held in the cognitive psychology literature. Namely, that the comprehension of a scenario involves constructing a mental representation of the described situation, which is aided by cues from the linguistic input [20][21][22][23][24].

For example, in comprehending the sentences "*The bottle fell off the shelf. The bottle smashed*", the main cognitive steps could be described as follows. First, the specific words in the first sentence reference distributional knowledge and cause a spotlight to fall on a neighbourhood of related terms in a high-dimensional distributional space [19][25][26]. Each of these spotlit terms then helps to prime relevant prior knowledge in long-term memory [27]. For example, knowledge relevant to the first sentence's situation may include that bottles are often fragile, that shelves are located at a height, that fragile things break when they hit the ground, etc. When the next sentence is read, the same procedure of distributional spotlighting and knowledge priming takes place. If any of the primed knowledge is used by inferences to connect the two sentences, then it will remain primed in case it is useful again; if the primed knowledge is not used, then it will be suppressed as irrelevant [28][29].

However, the amount of knowledge primed by a sentence pair depends on the size of the sentences' distributional spotlights. If sentences are close together in distributional space, then their spotlights will overlap almost completely and the overall coverage will be small. However, if the sentences are far apart, then their spotlights will fall on separate areas the overall coverage will be large (see Figure 1). The further apart the sentences are in distributional space, the greater the spotlight coverage and the more prior knowledge is primed. This means that distributionally distant sentence pairs have an advantage over distributionally close sentence pairs, because there is a greater chance that the knowledge required by the inference will already be primed. By retrieving prior knowledge (e.g. bottles are often fragile) and making the

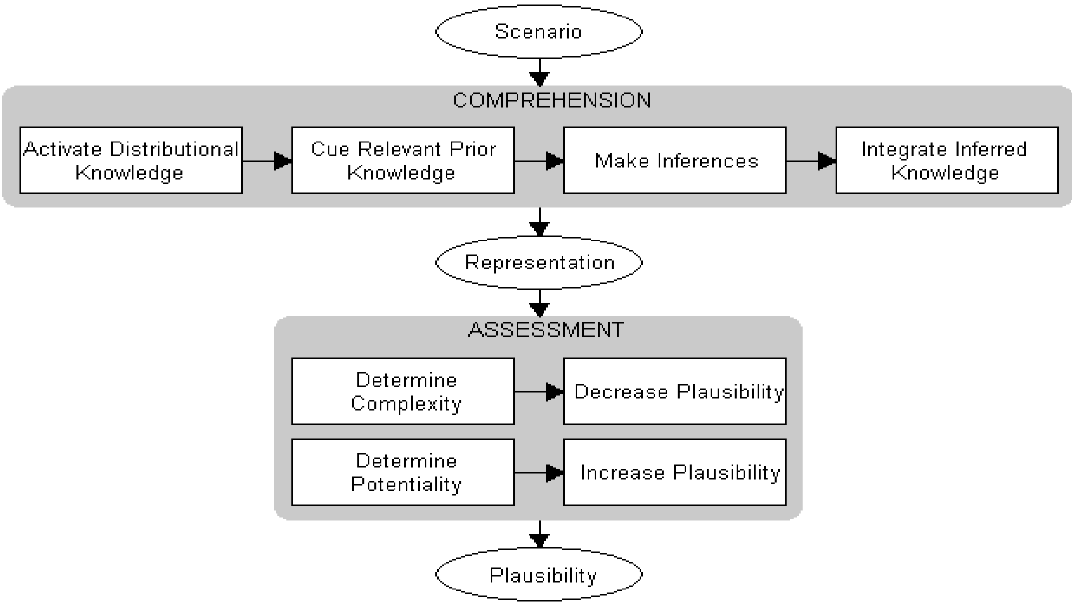

Figure 2. Diagram of processes in the Knowledge-Fitting Theory of Plausibility

relevant inferences (e.g. falling onto the floor *caused* the bottle to smash), the conceptual representation of the scenario is constructed and the sentence is said to be comprehended.

3.1.2 Assessment

Once a scenario has been comprehended, it must then undergo assessment to ascertain its plausibility. Assessment of a scenario involves examining the representation that has been produced by comprehension.

Plausibility assessment is dependent on two aspects of the representation; its *complexity* (how many inferences had to be made to connect events), and its *potentiality* (how much knowledge remains primed). For an example of the complexity aspect, with the sentences "*The bottle fell off the shelf. The bottle melted*" it might be possible to construct a representation where the bottle fell from the shelf onto the floor, which was made of metal, and which had somehow heated up enough to cause the bottle to melt. Clearly, this is not a very plausible scenario and it takes a lot of work to make it fit our knowledge about the world. The more inferences that need to be made to connect events in a scenario, and the more complex the representation grows, the less plausible the scenario becomes.

For an example of the potentiality aspect, with the sentence "*The bottle fell off the shelf. The bottle smashed*", the bottle smashed because it was fragile and hit the floor. Alternatively, the bottle may have smashed because it struck another shelf during its fall, or a table, or some other hard surface. Essentially, there are far more ways that a bottle can *smash* rather than *melt* – there are more ways that the scenario could have come about. When an individual is judging the plausibility of a scenario, it is unlikely that he or she will consider all of the possible versions. However, the possibility that more versions could potentially be constructed is reflected by the amount of knowledge that still remains primed. In both scenarios, the distributional spotlight of the first sentence ("*The bottle fell off the shelf*") primed knowledge relevant to the situation, such as that bottles are often fragile, that fragile things break when they hit the ground, etc. The melting scenario did not use this primed knowledge, and so it is suppressed. In assessing the scenario, the lack of primed knowledge suggests that there are no further versions of this scenario to be constructed and that there are no other ways in which we can make it fit our knowledge about the world. In contrast, the *smashing* scenario did make use of these examples of primed knowledge. The large amount of primed knowledge remaining suggests there are a further number of possible versions of the scenario that could be fleshed out if so desired, and suggests that there are many more ways that we could make this scenario fit our knowledge about the world. In short, the more knowledge that remains primed after the representation is built, the more potential versions of the scenario that could be constructed and the more plausible the scenario becomes.

3.2 Types of plausibility judgement

We have described how plausibility judgements are generally carried out, but we must also be aware that there are different ways in which one can judge plausibility. Recent work in [13][15][16] shows two different types of plausibility judgement task, which we shall discuss in turn.

The first type of plausibility judgement can be described as a binary plausibility decision, namely a decision of whether a scenario is plausible or not. To do this, one need only examine the complexity aspect of the representation during assessment (i.e. the number of inferences that had to be made to connect events). If the events could not be connected (e.g. no inferences could be made because of missing or contradictory information in prior knowledge), then the representation of the scenario is incomplete and the scenario will not be plausible. On the other hand, where a number of inferences can be seen to successfully connect the events, then the representation is complete and the scenario is plausible. For example, the statement that "*the balloon landed on a pin and burst*" is plausible because prior knowledge gives us the information that sharp things (the pin) can cause balloons to burst. However, saying that "*the balloon landed on a pin and melted*" is not plausible because there is little in our prior knowledge to suggest how a balloon could melt as the result of contact with a pin.

The second type of judgement is more involved than the first, requiring someone to ascertain exactly *how* plausible a scenario is. In order to do this, one must examine both aspects of the representation during assessment – complexity (how many inferences had to be made using prior knowledge) and potentiality (how much knowledge remains primed). As we noted earlier, the example where the bottle falls from the shelf and melts is not very plausible, because of the high complexity of several inferences being needed (such as the floor bring metal, and something heating it up to a high enough temperature). On the other hand, the example where the bottle falls from the shelf and smashes is quite plausible, because of the high potentiality of how primed knowledge could lead to other explanations (such as the bottle hitting the floor, or a table, or another shelf).

3.3 Location of effects

The two stages involved in making a plausibility judgement are quite distinct, and are supported by the empirical evidence.

Word-coherence affects only the comprehension phase (when words spotlight distributional knowledge). However, concept-coherence affects both the comprehension phase (when the representation of the scenario is built with reference to prior knowledge) and the assessment phase (when the structure of the scenario's representation is examined). As discussed above, the type of plausibility judgement task determines what aspects of the representation are assessed. In the case of *plausibility decision* times [13], only the complexity aspect of the representation need be examined, so there is not too much for the assessment phase to do. This makes the plausibility decision process more dependent on the comprehension phase, which is why both word- and concept-coherence effects are evident in decision times. In contrast, the task of *plausibility rating* [15][16] means that both the complexity and potentiality aspects of the representation must be examined, so there is a lot for the assessment phase to do. This makes the plausibility rating process more dependent on the assessment phase, which is why only the concept-coherence effect is evident in ratings, as it outweighs that of word-coherence.

4 PAM: PLAUSIBILITY ANALYSIS MODEL

PAM is a computational implementation of the Knowledge-Fitting theory of plausibility. PAM implements both the comprehension and assessment phases, incorporating knowledge of word-coherence and concept-coherence to provide judgements of plausibility that reflect those made by people [12][13][16]. The model takes sentence inputs and outputs an estimated plausibility decision time (in milliseconds) and a plausibility rating (from 0-10) for the scenario described in the sentences. PAM judges plausibility using a combination of commonsense reasoning (for concept-coherence) and distributional analysis (for word-coherence).

4.1 Comprehension phase

When a sentence is first read, each word helps to spotlight a certain area of distributional knowledge. PAM models this process by the use of a model of linguistic distributional knowledge, Latent Semantic Analysis (LSA: [19]). LSA (in the form used by PAM) is a statistical model of the distributional patterns of English words, which works by passing a window over a large corpus that represents the cumulative lifetime readings of an American first-year university student[1]. PAM uses LSA to calculate the 50 words that are the nearest neighbours of each sentence (i.e. the distributional spotlights), and then unifies the two sets of words. The number of unique words covered by the spotlights depends on how far apart the sentences are in LSA's distributional space. If the sentences are very close together then their spotlights will overlap completely, giving a count of 50 unique words. However, if the sentences are very far apart then their spotlights will be completely separate, giving a count of 100 unique words. This *distributional word count* represents the word-coherence factor, and is later used in the Assessment phase as a scaling variable in estimating the plausibility decision time and rating for the presented scenario. The higher the distributional word count, the more knowledge is primed, which means faster decision times and higher perceived plausibility.

As we have said, distributional information alone is not enough to form the basis of a judgement of plausibility; we also need a conceptual representation of the scenario. To do this, PAM breaks down each sentence into propositional form and makes the inferences between the sentences by fitting their propositions to information in the knowledge base.

For example, the sentence "*the pack saw the fox*" is represented as *see(pack, fox*). PAM must therefore check the predicate *see* in the knowledge base and determine if the arguments meet the conditions specified. The *see* predicate requires that its first argument is an animal (i.e. something must be an animal in order to see), and since the definition of pack shows that it contains dogs, and the type hierarchy for dog shows that it is an animal, the first condition of the see predicate is met. Also, the see predicate requires that its second argument is a non-abstract entity (i.e., something must be non-abstract in order to be seen). Since the type hierarchy of fox shows that it is an animal and not an abstract entity, the second condition of the see predicate is met. The way in which each condition is met is listed, and if all conditions are fulfilled, PAM returns this list as a path

When the first sentence has been represented, PAM moves onto processing the second sentence. The sentence "*the hounds growled*" is broken down into propositional form as *growl(hounds)* and PAM than searches for ways to meet the conditions of the *growl* predicate. The first condition is that the argument (hounds) must be an animal, which is easily met. However, there are other conditions as to *why* the hounds are growling, such as because they are generally aggressive, or because they are fighting amongst themselves. Some of these conditions lead to other predicates which have their own conditions attached, such as *hunt(hounds)* which requires that hounds must be predators and that the *fox* of the first sentence must be prey. It is likely that there are many different paths in the knowledge base that could be followed to fulfil the conditions of the *growl* predicate, and PAM will record them all. In this respect, PAM models group behaviour in plausibility judgement; rather than limit the representation to a single path that one individual may consider, PAM represents the set of paths that a group may consider and averages out the differences.

---

1 In LSA parlance, the analysis was done in the 'General Reading up to 1st Year College' semantic space, with pseudodoc comparison at maximum factors. In order to exclude misspellings and other very low frequency words, any words with a frequency in the corpus of less than 5 were excluded.

4.2 Assessment phase

When the comprehension phase is completed, it is the role of the assessment phase to analyse the structure of the path representation in order to estimate the plausibility decision time and to calculate the plausibility of the scenario.

PAM extracts three important variables from the representation:

1. Total Number of Paths P (the number of different ways the sentence conditions can be met in the knowledge base)
2. Mean Path Length L (the average count of how many different conditions must be met per path)
3. Proportion of "Hypothetical" Paths H (proportion of all paths that contain a condition that was only met by assuming the existence of something not explicitly mentioned – e.g. [*The bottle fell off the shelf. The bottle melted.*] is considered a plausible path if we assume a hypothetical furnace for the bottle may to fall into)

4.2.1 Plausibility decision times

In estimating the time needed to decide if a scenario is plausible or not, PAM uses the above variable L to calculate the concept-coherence of the scenario. A high path length (L) means a longer decision time, because more elaborate requirements must be met to verify the sentence. In addition, the comprehension time of a sentence is affected by syllable length, and to a lesser extent orthographic length, so PAM increases the decision time estimate as each of these increase.

Word-coherence has a strong effect on plausibility decision times, and PAM uses the distributional word count (calculated in the Comprehension phase) to model this. If the two sentences in the pair are very far apart, then they will have the maximum distributional word count of 100. However, if the two sentences are very close to each other, then they will have the minimum distributional word count of 50. The higher the distributional word count, the more knowledge that the sentences prime, and the faster the plausibility decision time becomes. PAM therefore uses the distributional word count to scale down the estimated response time.

It has been demonstrated that using this approach allows PAM to produce estimates of plausibility decision times that correlate highly with human responses (r=0.633, r^2=0.401, p<0.0001, N=60) [12]. The human data modelled was taken from [13], and is graphed against PAM's output in Figure 3.

4.2.2 Plausibility ratings

PAM uses the three variables above to calculate the concept-coherence of the scenario, and return a rating between 0 (not plausible) and 10 (completely plausible). In short, a high number of paths (P) means higher plausibility, because there are more possible ways that the scenario can be represented. A high mean path length (L) means lower plausibility, because elaborate requirements must be met to verify the sentence. Finally, a high proportion of hypothetical paths (H) means lower plausibility, because it assumes the existence of entities that may not be present.

When the path rating has been computed, PAM applies word-coherence (the distributional word count calculated in the comprehension phase) as a scaling variable. The magnitude of this scaling is less than that of other variables, but still has a perceptible effect. In this way, PAM models the small difference in plausibility ratings found between versions of sentence pairs that vary in their distributional distance but are conceptually identical (see pairs i and ii).

Table 1. Mean Plausibility ratings per inference type from human raters and from PAM.

Inference Type	Human Rating	PAM Rating
Causal	7.8	7.9
Attributal	5.5	5.7
Temporal	4.2	5.0
Unrelated	2.0	0.9

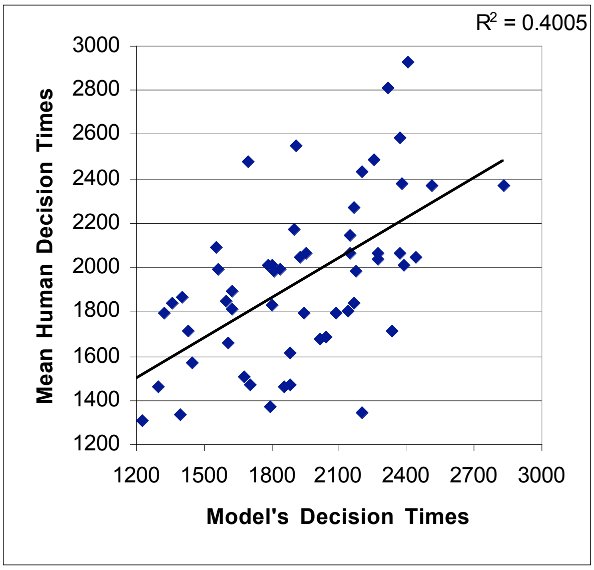

Figure 3. PAM's performance against human responses in modelling plausibility decision times

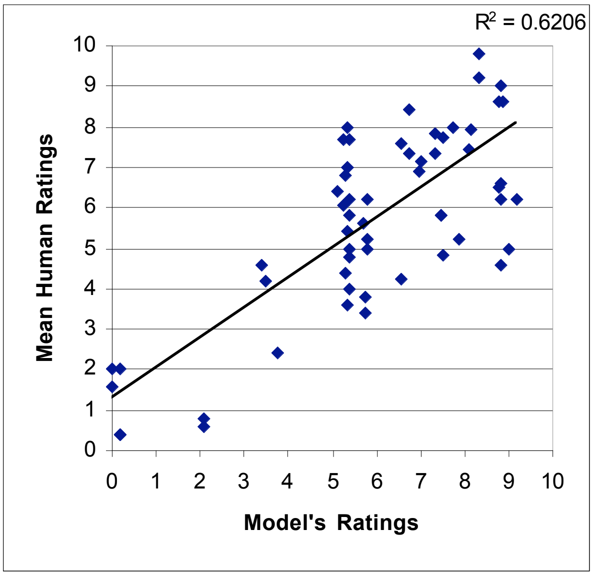

Figure 4. PAM's performance against human responses in modelling plausibility ratings

It has been demonstrated that using this approach, PAM can produce plausibility ratings that correlate highly with human plausibility judgements (r=0.788, r^2=0.621, p<0.0001, N=60) [12]. The human data modelled was taken from [16], and is graphed against PAM's output in Figure 4. Additionally, Table 1 illustrates mean ratings for scenarios that invite different types of bridging inference, comparing those produced by people to those ratings produced by PAM.

5 DISCUSSION

In this paper we have presented the new Knowledge-Fitting Theory of Plausibility and the Plausibility Analysis Model (PAM), which is the first account to specifically and accurately address human plausibility judgements. Our theory of plausibility can explain the some counterintuitive empirical findings regarding word- and concept-coherence, and provides simulations for testing in further studies.

PAM is the computational implementation of the Knowledge-Fitting theory, and as such is the first cognitive model of human plausibility judgements. The importance of word- and concept-coherence in people's plausibility judgements is clear, and in a novel paradigm, we have integrated both these factors in our model. Future work in the field of plausibility must also take account of both distributional knowledge and conceptual prior knowledge, as well as the interactions between them.

ACKNOWLEDGMENTS

This work has been part-funded by a grant from the Irish Research Council for Science, Engineering and Technology.

REFERENCES

[1] P. Lemaire and M. Fayol. When plausibility judgments supersede fact retrieval: The example of the odd-even rule effect in simple arithmetic. *Memory and Cognition, 23*, 34-48, (1995).

[2] L. M. Reder. Plausibility judgments vs. fact retrieval: Alternative strategies for sentence verification. *Psychological Review, 89*(3), 250-280, (1982).

[3] L. M. Reder and B. H. Ross. Integrated knowledge in different tasks: The role of retrieval strategy on fan effects. *Journal of Experimental Psychology: Learning, Memory and Cognition, 9*, 55-72, (1983).

[4] L. M. Reder, C. Wible and J. Martin. Differential memory changes with age: Exact retrieval versus plausible inference. *Journal of Experimental Psychology: Learning, Memory and Cognition, 12*(1), 72-81, (1986).

[5] M. J. Pickering and M. J. Traxler. Plausibility and recovery from garden paths: An eye-tracking study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 940-961, (1998).

[6] S. R. Speer and C. Clifton. Plausibility and argument structure in sentence comprehension. *Memory and Cognition, 26*(5), 965-978, (1998).

[7] M. J. Traxler and M. J. Pickering. Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language, 35*, 454 - 475, (1996)

[8] A. Collins and R. Michalski. The logic of plausible reasoning: A core theory. *Cognitive Science, 13,* 1-49, (1989).

[9] E. E. Smith, E. Shafir and D. Osherson. Similarity, plausibility, and judgments of probability. *Cognition, 49*, 67-96, (1993).

[10] F. Costello and M. T. Keane. Efficient Creativity: Constraints on conceptual combination. *Cognitive Science, 24,* 299-349, (2000).

[11] F. Costello and M. T. Keane. Alignment versus diagnosticity in the comprehension and production of combined concepts. *Journal of Experimental Psychology: Learning, Memory & Cognition, 27*, 255-271, (2001).

[12] L. Connell and M. T. Keane. PAM: A Cognitive Model of Plausibility. *Proceedings of the Twenty-Fifth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum, (2003).

[13] L. Connell and M. T. Keane. The knowledge-fitting theory of plausibility. *Manuscript in preparation.*

[14] A. Black, P. Freeman and P. N. Johnson-Laird. Plausibility and the comprehension of text. *British Journal of Psychology, 77*, 51-60, (1986).

[15] L. Connell and M. T. Keane. The roots of plausibility: The role of coherence and distributional knowledge in plausibility judgements. *Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society* (p. 998). Hillsdale, NJ: Erlbaum, (2002).

[16] L. Connell and M. T. Keane. What plausibly affects plausibility? Concept-coherence and distributional word-coherence as factors influencing plausibility judgements. *Manuscript in submission.*

[17] P. Thagard. Coherence in thought and action. Cambridge, MA: MIT Press, 2000.

[18] M. Lapata, S. McDonald and F. Keller. Determinants of adjective-noun plausibility. *Proceedings of the 9th Conference of the European Chapter of the Association for Computational Linguistics*, (pp. 30-36). San Mateo, CA: Morgan Kaufmann, (1999).

[19] T. K. Landauer and S. T. Dumais. A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge. *Psychological Review, 104*, 211-240, (1997).

[20] M. A. Gernsbacher. Language comprehension as structure building. Hillsdale, NJ: Erlbaum, (1990).

[21] M. Singer, A. C. Graesser and T. Trabasso. Minimal or global inferences during reading. *Journal of Memory and Language*, 33, 421-441, (1994).

[22] P. N. Johnson-Laird. *Mental Models*. Cambridge: Cambridge University Press, 1983.

[23] G. McKoon and R. Ratcliff. Inference during reading. Psychological Review, 99, 440-466, (1992).

[24] R. A. Zwaan and G. A. Radvansky. Situation models in language comprehension and memory. *Psychological Bulletin*, 123, 162-185, (1998).

[25] C. Burgess, K. Livesay and K. Lund. Explorations in context space: Words, sentences, discourse. *Discourse Processes, 25*, 211-257, (1998).

[26] K. Lund, C. Burgess and R. A. Atchley. Semantic and associative priming in high-dimensional semantic space. *Proceedings of the 17th Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum, (1995).

[27] M. Halldorson and M. Singer. Inference processes: integrating relevant knowledge and text information. *Discourse Processes, 34*, 145-161, (2002).

[28] M. A. Gernsbacher and M. Faust. The role of suppression in sentence comprehension. In G.B. Simpson (Ed.), *Understanding word and sentence.* Amsterdam: North Holland, 1991.

[29] M. A. Gernsbacher and R. W. Robertson. Reading skill and suppression revisited. *Psychological Science, 6*, 165-169, (1995).