Acta Psychologica 160 (2015) 194–201

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/ locate/actpsy

Conditional sentences create a blind spot in theory of mind during narrative comprehension☆

Matthew Haigh [a,*], Jean-François Bonnefon [b]

[a] Department of Psychology, Northumbria University, United Kingdom
[b] Toulouse School of Economics, Toulouse, France

ARTICLE INFO

Article history:
Received 12 September 2014
Received in revised form 21 May 2015
Accepted 25 May 2015
Available online 2 June 2015

Keywords:
Conditionals
Theory of Mind
Desire
Discourse comprehension
Eye tracking

ABSTRACT

We identify a blind spot in the early Theory of Mind processing of conditional sentences that describe a protagonist's potential action, and its predictable consequences. We propose that such sentences create expectations through two independent channels. A decision theoretic channel creates an expectation that the action will be taken (viz., not taken) if it has desirable (viz., undesirable) consequences, but a structural channel acts in parallel to create an expectation that the action will be taken, irrespective of desirability. Accordingly, reading should be disrupted when a protagonist avoids an action with desirable consequences, but reading should not be disrupted when a protagonist takes an action with undesirable consequences. This prediction was supported by the eye movements of participants reading systematically varied vignettes. Reading was always disrupted when the protagonist avoided an action with desirable consequences, but disruptions were either delayed (Experiment 1) or recovered from faster (Experiment 2) when the protagonist took an action with undesirable consequences.

1. Introduction

In order to predict the actions of other people as a narrative or a social situation unfolds, it is necessary to keep track of their beliefs and their desires through online, fast *Theory of Mind* inferences (Kovács, Téglás, & Endress, 2010). Recent studies have shown that the desires of others are consistently at the forefront of an observer's mind (Malle & Holbrook, 2012), even when these desires are complex and conflicting (Ferguson & Breheny, 2011). Avoidance desires, though, seem to pose a problem of their own. An individual has an avoidance desire when he or she wants to avoid some state of the world, rather than to see it realised. Children, adults, and older adults alike find it difficult to track avoidance desires (Apperly, Warren, Andrews, Grant, & Todd, 2011; German & Hehman, 2006; Leslie, German, & Pollizi, 2005). Furthermore, they appear to rely on slow and deliberate mental processes to track avoidance desires, as if fast and automatic processes could not be relied on in these circumstances.

In this article, we employ eye tracking measures of narrative comprehension to show that avoidance desires hide in a blind spot of our fast, automatic Theory of Mind (cf. Haigh & Bonnefon, 2015, for an application of this method to approach desires). We consider narratives that describe the potential action (P) of a protagonist, as well as its predictable consequences (Q), under the form of a conditional sentence ("if P, then Q"). The literature we review suggests that these sentences create expectations through two separate channels, whose output can diverge. This divergence can in turn lead readers to expect actions that contradict the avoidance desire of the protagonist. We test this model in two eye tracking experiments that recorded various measures of reading fluency for systematic variations of the narrative.

1.1. The decision-theoretic channel

Let us consider a conditional sentence describing the potential action of a protagonist, as well as the desirable (1a) or undesirable (1b) consequences of that action:

(1a) Jayne said to Robert "If I have oysters for my main course, I'll be a very happy lady";
(1b) Jayne said to Robert "If I have oysters for my main course, I will be very ill".

What do we expect Jayne to do? Most people predict that Jayne will do what serves her interest best (Bonnefon, 2009, 2012; Bonnefon & Hilton, 2004; Bonnefon & Sloman, 2013). That is, a majority of people who read 1a predict that Jayne will have the oysters — and a majority of people who read 1b predict that Jayne will avoid the oysters.

☆ This work was supported by an Experimental Psychology Society (EPS) Small Grant awarded to the first author. The EPS had no role in study design, the collection, analysis and interpretation of data, the writing of the report or the decision to submit the paper for publication.

* Corresponding author at: Department of Psychology, Northumbria University, Newcastle upon Tyne, NE1 8ST, United Kingdom.
E-mail address: matthew.haigh@northumbria.ac.uk (M. Haigh).

http://dx.doi.org/10.1016/j.actpsy.2015.05.009
0001-6918/

In other words, a "utility" conditional (Bonnefon, 2009) such as 1a or 1b creates an expectation that the protagonist will make the antecedent true when its consequent is desirable (if P then desirable Q; therefore, protagonist will do P), and it creates an expectation that the protagonist will make the antecedent false when its consequent is undesirable (if P then undesirable Q; therefore, protagonist will do not-P). Readers expect that protagonists will take actions that lead to positive consequences, and not take actions that lead to negative consequences. These expectations are rational if we assume that others will act in a self-interested manner, by taking actions that increase their personal utility and avoiding actions that would decrease their personal utility (Bonnefon, 2009).

This view is what we call the *decision-theoretic* channel: conditional sentences create expectations about the truth or falsity of the antecedent based on the desirability of its consequences. Expectations delivered by the decision-theoretic channel are a textbook example of Theory of Mind inferences: Based on the mental states assigned to the protagonist (a desire to attain or avoid the consequent, and a belief that the antecedent will realise or prevent the consequent), an inference is derived about whether or not the protagonist has the intention to perform the action described in the antecedent. Conditional sentences, however, can create expectations based on another channel, independent of Theory of Mind, that we call the *structural channel*.

1.2. The structural channel

In parallel to reasoning tasks featuring utility conditionals (with desirable or undesirable contents), other studies focused on the Event Related Potentials (ERPs) associated with reading neutral-content conditionals (Bonnefond & Van der Henst, 2009, 2013). One key aspect of these studies is that they compared conditionals followed by a sentence that matched the antecedent (as in 2a), to conditionals followed by sentences that mismatched the antecedent (as in 2b):

(2a) If John is sleeping, then he is snoring; John is sleeping;
(2b) If John is sleeping, then he is snoring; John is singing.

These ERP studies provide consistent evidence that people who read a conditional "if P then Q" expect that P is true. Reading that P is true as in 2a prompts a P3b component (which is typically associated with expected stimuli) whereas reading about a different action (as in 2b) prompts a N2 component (which is typically associated to unexpected stimuli). This is more than just a simple (mis)match effect as readers do not expect to be told that Q is true (John is snoring), even though it is explicitly mentioned in the conditional rule (Bonnefond et al., 2012).

These findings suggest that P may be the fast acting, default expectation following a conditional of the form if P then Q; a claim that is supported by most psychological theories of conditionals (e.g., Evans & Over, 2004; Johnson-Laird & Byrne, 2002; Sperber & Wilson, 1995). These theories share the assumption that the first step of representing a conditional statement is to assume that the antecedent (P) is true. Note that this expectation seems independent of the desirability of P and Q, which played a critical role in the decision-theoretic channel. It appears that the mere structure of the conditional triggers expectations that the antecedent must be true. From a decision theoretic perspective, this expectation is rational when the consequences of an action (P) have neutral or positive utility, but may be not be rational where the action has consequences with negative utility (e.g., assuming that Jayne will have the oysters even though they will make her ill).

This is what we call the structural channel: conditional sentences create uniform expectations about the truth of the antecedent, irrespective of desirability.

1.3. Two channels make for a blind spot in Theory of Mind

We propose that conditional sentences create online expectations based on two parallel channels. When people read a conditional sentence "if P then Q", they form an expectation that P is true through the structural channel, and they also form a decision theoretic expectation that P is true (viz., false) when Q is desirable (viz., undesirable). What are the consequences for online narrative comprehension? Let us look at four combinations of sentences:

1. "If I have oysters for my main course, I'll be a very happy lady". With this in mind she decided to *order* the oyster special.
   After reading the conditional, the structural channel creates an expectation that Jayne will have oysters, and the decision-theoretic channel does the same. Accordingly, readers should have no problem processing a subsequent region of text that describes Jayne ordering the oyster special.
2. If I have oysters for my main course, I'll be a very happy lady". With this in mind she decided to *avoid* the oyster special.
   The structural channel creates an expectation that Jayne will have oysters, and the decision-theoretic channel does the same. The subsequent description of Jayne avoiding the oyster special is therefore in violation of the expectations created by both channels, and should thus create substantial reading disruption.
3. "If I have oysters for my main course, I will be very ill". With this in mind she decided to *order* the oyster special.
   This is the combination where Theory of Mind may be susceptible to a blind spot. The structural channel creates an expectation that Jayne will have the oysters, but the decision-theoretic channel creates an expectation that she won't. This means that readers must simultaneously anticipate P and not-P, thus not precluding any action. Therefore, reading disruption may be minimal even though the protagonist is taking an ostensibly irrational action.
4. "If I have oysters for my main course, I will be very ill". With this in mind she decided to *avoid* the oyster special.
   The structural channel creates an expectation that Jayne will have the oysters, but the decision-theoretic channel creates an expectation that she won't. Again the reader must simultaneously anticipate P and not-P. Because the subsequent sentence describes a rational action, reading should not be disrupted.

Predictions based on parallel structural and decision theoretic channels have implications for how we understand online Theory of Mind processing. Compare the second and third vignettes, which both describe the protagonist acting against her best interests. In the second vignette (which implies an approach desire), we expect readers to detect that the character acted against her best interest because her action contradicts the expectations of both the structural and decision theoretic channels. However, in the third vignette (which implies the avoidance desire of not getting ill) the parallel channels build opposite expectations (P and not-P), thus not precluding any action. As a result, people may not easily detect that the character acted against her best interest. This essentially creates a Theory of Mind blind spot for avoidance desires during online processing, consistent with previous findings that suggest avoidance desires are especially difficult to process.

To test our prediction, we created narrative vignettes similar to the Jayne example, in which the decision theoretic and structural channels produced either matching or conflicting predictions about the truth of the antecedent. To reveal how these channels influenced online processing, we analysed the eye movements of participants as they read follow up sentences to the conditional, that either asserted the truth (P) or the falsity (not-P) of its antecedent.

2. Experiment 1

Participants were presented with a series of vignettes containing conditional sentences describing a protagonist's potential action (P) and

its consequences (Q). These conditionals were manipulated to have either a desirable or undesirable consequence for the protagonist. A sentence followed in which the protagonist either took action P or did not take action P (not-P). These variables were fully crossed to create four experimental conditions. The dependent variables were various eye movement measures associated with reading two specific regions of text. The critical analysis region was the region of text describing whether the action was taken or not taken (see Fig. 1). We also analysed the following, post-critical, sentence to capture any late onset or residual effects. In line with the predictions outlined above, we expect that reading will be disrupted when the protagonist avoids an action with desirable consequences (relative to a condition in which the protagonist avoids an action with undesirable consequences) but reading should not be disrupted when the protagonist takes an action with undesirable consequences (relative to a condition in which the protagonist takes an action with desirable consequences).

2.1. Participants

Thirty two psychology undergraduates (20 females, mean age 21.2 years) from Northumbria University volunteered in exchange for course credit. Testing with one additional participant was terminated due to a software malfunction. All were native English speakers with no language impairment and normal or corrected-to-normal vision.

2.2. Design & materials

Each participant read a series of vignettes containing conditionals that described a possible action P (e.g., if I have oysters…') and its predicted consequence, Q. The consequence was manipulated so that it described either a desirable or undesirable outcome for the speaker (e.g., if I have oysters I'll be a very happy lady/if I have oysters I'll be very ill). The next sentence described an action by the speaker that either affirmed (P) or denied (not-P) the antecedent P (e.g., the agent orders/avoids the oysters). This fully crossed 2 × 2 (Consequent Utility × Protagonist Decision) repeated measures design gives four conditions.

Each vignette consisted of three sentences. Sentence one introduced two fictional characters and contained a conditional statement uttered by the first. Sentence two then described the subsequent actions of the first character. Sentence three provided a neutral continuation that was identical within items across all four conditions.

Ninety six vignettes were constructed (24 items, each with four versions; see supplementary material for full list of items). One version of each item was placed into one of four Latin squared presentation lists. Each participant read one list containing 24 vignettes (eight participants read each list) interspersed with 21 unrelated filler vignettes. The order of presentation was randomised for each participant. Comprehension questions followed half of the vignettes. No feedback was given on accuracy. Mean comprehension accuracy was 82%.

2.3. Procedure

Eye movements were recorded using an EyeLink 1000 eye tracker in the Desktop Mount configuration. Viewing was binocular and recordings were made from the right eye at 1000 Hz. Vignettes were presented in size 22 Ariel font on a CRT monitor placed 80 cm from the participants' eyes. A chin and forehead rest was used to minimise head movement.

The eye tracker was calibrated at the beginning of the experiment using nine fixation points and this procedure was repeated as necessary. Before each trial a gaze trigger appeared in the top left quadrant of the screen where the first word would be displayed. After fixating on the gaze trigger a vignette appeared and the participant read this silently at their normal rate for comprehension. After reading the vignette a button on a handheld controller was pressed to progress. Following one half of the vignettes a comprehension question appeared and the participant answered 'yes' or 'no' by pressing the corresponding button on the controller.

2.4. Analysis

Each vignette was split into two analysis regions, as illustrated in Fig. 1. The critical region was the final verb phrase in sentence two, that described the decision of the protagonist to take the action (P) or not to take it (not-P). We also analysed the post-critical sentence to capture any late onset or residual effects.

We report data from four measures of eye movement. Two of these were chronometric, measuring time taken to read a region of text and two measured regressive eye movements into and out of a given region.

Regressions Path time (ms): Total time from first entering a region of text to first exiting it to the right (including time spent outside of the region to the left). This measure provides information about the time taken to go past a region of text after it is first entered. If a region is not exited to the right (e.g., on the final sentence) then Regression Path measures the time from first entering the region of text until the participant presses a button to indicate they have finished reading.

Total Time (ms): The sum of all fixation durations in a region.

First Pass Regressions Out (FPRO): Percentage of trials in which one or more first pass fixations in a region were followed by a fixation to an earlier point in the vignette. This indicates the degree to which normal left to right eye movements are disrupted when first reading a region of text.

Regressions In: Percentage of trials where one or more fixations in a region were preceded by a fixation to a later point in the vignette. This provides information about how frequently a region of text was re-visited.

	Takes action P	Does not take action P
"if P then desirable Q"	Jayne said to Robert, "if I have oysters for my main course, I'll be a very happy lady". With this in mind she decided to |CRITICAL order the oyster special. |POST-CRITICAL The waiter quickly noted down their requests and headed straight to the kitchen.|	Jayne said to Robert, "if I have oysters for my main course, I'll be a very happy lady". With this in mind she decided to |CRITICAL avoid the oyster special. |POST-CRITICAL The waiter quickly noted down their requests and headed straight to the kitchen. |
"if P then undesirable Q"	Jayne said to Robert, "if I have oysters for my main course, I will be very ill". With this in mind she decided to |CRITICAL order the oyster special. |POST-CRITICAL The waiter quickly noted down their requests and headed straight to the kitchen. |	Jayne said to Robert, "if I have oysters for my main course, I will be very ill". With this in mind she decided to |CRITICAL avoid the oyster special. |POST-CRITICAL The waiter quickly noted down their requests and headed straight to the kitchen. |
	Q: Were oysters on the menu? NO	YES

Fig. 1. Example of the four experimental conditions and analysis regions in Experiment 1.

Fixations of less than 80 ms were pooled with adjacent fixations, while fixations shorter than 40 ms were excluded if they were not within three characters of another fixation. All fixations greater than 1200 ms were truncated.

A series of 2 × 2 (Consequent utility × Protagonist Decision) repeated measures ANOVAs were performed with participants (F_1) and items (F_2) as random factors.

2.5. Results & discussion

Table 1 displays the means and standard errors of our five measures for all conditions and analysis regions. Table 2 displays the ANOVA results by participants and by subjects. All main effects of Protagonist Decision on the critical region are influenced by unavoidable lexical differences between conditions and will not be discussed further. We designed the experiment to focus on interaction effects, driven by differences between conditions to lexically identical regions of text. Specifically, we are interested in planned comparisons between the two conditions in which the protagonist takes action P and planned comparisons between the two conditions in which the protagonist does not take action P (i.e., not-P).

Table 2 shows several significant (and marginally significant) interaction effects for various reading time and eye movement measures on the critical and post-critical analysis regions. These interaction effects are interpreted using paired t-tests with Bonferroni corrected p values (reported in Table 3). On the critical region we found interaction effects relating to measures of Regression Path time and Total Time. These interactions all shared the same asymmetrical pattern. When the critical sentence asserted P ('order the oyster special') this region was read without disruption, regardless of how (un)desirable its consequences would be. When the critical sentence asserted not-P ('avoid the oyster special') it was read without disruption when P would lead to an undesirable consequence, but caused significant disruption when P would lead to a desirable consequence. This pattern is entirely consistent with our hypothesis that a structural channel operates alongside a decision theoretic channel, at least during the processing of the critical region.

Residual interaction effects were observed on First Pass Regressions Out of the post-critical region and also on Regression Path time, but these interactions were more symmetrical. On First Pass Regressions Out neither of the paired comparisons was statistically significant but the percentage point difference between the two pairs of means was numerically similar in the 'P' and 'not-P' conditions. For Regression Path time the difference between the two 'not-P' conditions was significant (p1 = .014; p2 = .004) and the difference between the two conditions in which 'P' was asserted came close to significance (p1 = .064; p2 = .076). The pattern of means on this later region of text is more in line with a purely decision theoretic interpretation of the protagonist's behaviour.

In line with our predicted Theory of Mind blind spot, eye movements to the critical region revealed disrupted reading the when a protagonist avoided an action that would bring positive consequences (relative to a condition in which the protagonist avoided an action that would bring negative consequences), but no such disruption when the protagonist took an action with negative consequences (relative to a condition in which the protagonist took an action with positive consequences). However, on the post-critical region there was evidence of delayed disruption in this latter condition. These findings are consistent with our suggestion that the structural channel creates an expectation that P will be true when reading a conditional 'if P then Q', independently of the desirability of Q for the protagonist. When P is undesirable, this structural expectation initially masks the decision-theoretic expectation that action P will not be taken, creating a temporary blind spot in online Theory of Mind processing.

One aspect of our materials, though, requires closer scrutiny. Our vignettes used various precursors for the critical sentence describing the action of the protagonist, and it is possible that some of these precursors may have made the critical sentence less surprising than others. Most sentences began straight away with "He then…" or "She then…", whereas other sentences began with "With this in mind…" or "This prompted her to…". While beginnings like "He then…" or "She then…" appear fairly neutral, a precursor like "With this in mind…" may suggest that the protagonist had other consequences in mind (positive or negative), which may have compensated the positive or negative consequences explicitly mentioned in the conditional. In contrast, a precursor like "This prompted her to…" would maximise the surprise created by the protagonist's decision by emphasising the contingency between the implied desire and their subsequent action.

As a stronger test of our hypothesis, we conducted an additional experiment in which we drew more attention to the contingency between a protagonist's implied approach or avoidance desire and their subsequent action. This should maximise the likelihood of readers noticing decision theoretic anomalies that might otherwise be hidden in a Theory of Mind blind spot. We achieved this by always introducing the critical sentence using the precursor "This prompted her to …". Such a precursor creates a strong contingency between the conditional and the subsequent action and should therefore maximise the surprising nature of the critical sentence. We also presented a comprehension question after all of our items (rather than 50%) to encourage a deeper level of engagement. The aim of Experiment 2 was to determine whether there would still be evidence of a Theory of Mind blind spot even when the vignettes emphasised a direct link between a protagonist's desire and their subsequent action.

Table 1
Experiment 1 mean reading times and regressions by analysis region and condition (Means averaged over subjects, standard errors in parentheses).

Utility of *q*/protagonist decision	First pass Regressions out (%)	Regression path (ms)	Regressions In (%)	Total Time (ms)
Critical region				
1. Desirable/P	17.5 (2.7)	948 (50)	13.3 (3.5)	952 (47)
2. Desirable/not-P	23.7 (3.5)	1234 (64)	20.0 (3.1)	1225 (57)
3. Undesirable/P	15.8 (3.4)	949 (68)	11.8 (3.0)	961 (58)
4. Undesirable/not-P	11.6 (2.1)	1009 (46)	15.4 (3.1)	1084 (52)
Post-critical region [a]				
1. Desirable/P	51.3 (5.9)	2743 (243)	n/a	2013 (88)
2. Desirable/not-P	61.2 (6.0)	3312 (271)	n/a	2069 (90)
3. Undesirable/P	59.6 (5.2)	3003 (260)	n/a	2083 (90)
4. Undesirable/not-P	55.2 (4.8)	2719 (167)	n/a	2002 (84)

[a] Regressions In were not applicable to post-critical region as this was always the final sentence of the vignette.

3. Experiment 2

3.1. Participants

Twenty eight Northumbria University students (18 females, mean age 24.1 years) volunteered in exchange for £6 cash. An additional two participants did not complete the experiment due to poor calibration and their data were excluded. All were native English speakers with no language impairment and normal or corrected-to-normal vision.

3.2. Design & materials

The design was identical to Experiment 1. Two changes were made to the materials used in Experiment 1. First, the beginning of sentence two was re-worded so that it created a strong contingency between the conditional and the subsequent decision. This was achieved by using the precursor 'This prompted him/her to…' in all of our 24 items (see Fig. 2). The new precursor necessitated some minor changes to

Table 2
Experiment 1 ANOVA results for each region and relevant measure.

Region	Measure	Predictor	By participants			By items		
			$F_1(1,31)$	p	η_p^2	$F_2(1,23)$	p	η_p^2
Critical	Regression path	Utility	6.46	.016*	.172	7.22	.013*	.239
		Decision	24.52	<.001*	.442	4.81	.039*	.173
		Interaction	4.31	.046*	.122	5.06	.034*	.180
	Total Time	Utility	4.21	.049*	.120	5.87	.024*	.203
		Decision	27.27	<.001*	.468	10.79	.003*	.319
		Interaction	3.77	.061	.108	6.65	.017*	.224
	FPRO	Utility	8.39	.007*	.213	15.78	<.001*	.407
		Decision	<1			<1		
		Interaction	3.3	.079	.096	4.20	.052	.154
	Regressions In	Utility	<1			1.32	.262	.054
		Decision	2.85	.101	.084	2.59	.121	.101
		Interaction	<1			<1		
Post-critical	Regression path	Utility	3.18	.085	.093	3.90	.060	.145
		Decision	<1			2.16	.155	.086
		Interaction	9.37	.005*	.232	12.30	.002*	.348
	Total Time	Utility	<1			<1		
		Decision	<1			<1		
		Interaction	1.70	.202	.052	.243	.133	.096
	FPRO	Utility	<1			<1		
		Decision	<1			<1		
		Interaction	5.76	.023*	.157	4.42	.047*	.161

* $p < .05$

the subsequent text for some items. Second, comprehension questions were presented after all experimental and filler items, rather than after half of the items (mean comprehension accuracy in Experiment 2 was 88%). Each participant read one Latin squared list containing 24 vignettes (seven participants read each list) interspersed with the same 21 filler vignettes used in Experiment 1.

3.3. Procedure

The procedure was identical to Experiment 1.

3.4. Analysis

The analysis regions were defined in the same way as Experiment 1 (see Fig. 2). A series of 2 × 2 (Consequent utility × Protagonist Decision) repeated measures ANOVAs were performed with participants (F_1) and items (F_2) as random factors.

3.5. Results & discussion

Table 4 shows the means and standard errors of our five measures for all conditions and analysis regions. Table 5 shows several significant interaction effects for various reading time and eye movement measures on the critical and post-critical regions. These interaction effects are interpreted using paired t-tests with Bonferroni corrected p values (reported in Table 6). On the critical region we found significant interaction effects relating to measures of Regression Path time, Total Time, and First Pass Regressions Out. These interactions shared the same symmetrical pattern. When the critical region asserted P ('order the oyster special') it was read without disruption when the protagonist

Table 3
Paired comparisons between lexically identical regions of text following significant and marginal interaction effects in Experiment 1.

Region	Measure	Paired comparison	Mean difference	By participants			By items		
				t_1 (31)	p	d	t_2 (23)	p	d
Critical	Regression path	Desirable/P Undesirable/P	1 ms	.02	1.00	.01	.06	1.00	.01
		Desirable/not-p Undesirable/not-p	225 ms	3.04	.010*	.55	3.07	.010*	.73
	Total Time	Desirable/P Undesirable/P	9 ms	.21	1.00	.04	.34	1.00	.01
		Desirable/not-p Undesirable/not-p	141 ms	2.42	.044*	.43	3.19	.008*	.65
	FPRO	Desirable/P Undesirable/P	1.7 pp	.44	1.00	.08	.52	1.00	.10
		Desirable/not-p Undesirable/not-p	12.1 pp	3.35	.004*	.62	4.50	<.001*	1.15
Post-critical	Regression path	Desirable/P Undesirable/P	259 ms	2.25	.064	.40	2.20	.076	.45
		Desirable/not-p Undesirable/not-p	593 ms	2.86	.014*	.57	3.42	.004*	.72
	FPRO	Desirable/P Undesirable/P	8.3 pp	1.91	.130	.34	1.50	.294	.32
		Desirable/not-p Undesirable/not-p	6 pp	1.27	.426	.23	1.27	.436	.26

* Bonferroni corrected p value < .05.

	Takes action P	Does not take action P
"if P then desirable Q"	Jayne said to Robert, "if I have oysters for my main course, I'll be a very happy lady". This prompted her to |CRITICAL order the oyster special. |POST-CRITICAL The waiter quickly noted down their requests and headed straight to the kitchen.|	Jayne said to Robert, "if I have oysters for my main course, I'll be a very happy lady". This prompted her to | CRITICAL avoid the oyster special. | POST-CRITICAL The waiter quickly noted down their requests and headed straight to the kitchen. |
"if P then undesirable Q"	Jayne said to Robert, "if I have oysters for my main course, I will be very ill". This prompted her to | CRITICAL order the oyster special. | POST-CRITICAL The waiter quickly noted down their requests and headed straight to the kitchen. |	Jayne said to Robert, "if I have oysters for my main course, I will be very ill". This prompted her to | CRITICAL avoid the oyster special. | POST-CRITICAL The waiter quickly noted down their requests and headed straight to the kitchen. |
	Q: Were oysters on the menu? NO	YES

Fig. 2. Example of the four experimental conditions and analysis regions in Experiment 2.

desired the consequences of this decision. However, there was relative disruption to reading when this same region of text was presented in a context where the decision would lead to an undesirable consequence. Likewise, when the critical region asserted not-P ('avoid the oyster special') it was read without disruption when the consequences of P were undesirable, but resulted in relative disruption when the consequences of P were actually desirable. This pattern of data suggests that processing of critical region was guided solely by the decision theoretic channel, with no evidence of an initial Theory of Mind blind spot.

On the post-critical region we again found interaction effects, this time associated with Regression Path time and First Pass Regressions Out. Importantly, Table 6 shows that these interactions were asymmetrical. When sentence two asserted P ('order the oyster special') the post-critical region (sentence three) was read without disruption, regardless of how (un)desirable its consequences would be. When sentence two asserted not-P ('avoid the oyster special') the post-critical region was read without disruption when P would lead to an undesirable consequence, but caused significant disruption when P would lead to a desirable consequence. The same pattern of means was also found for the percentage of trials with a regressive saccade from the post-critical region back in to the critical region. This pattern of means is identical to that found on the critical region of Experiment 1. These results indicate that the structural channel helps readers to more easily recover from the disruption caused by a decision theoretic anomaly.

Experiment 2 was designed to draw attention to the contingency between the protagonist's desires and their subsequent action. The results of this experiment show that emphasising this direct link causes the blind spot to disappear on the critical region (i.e., reading was initially disrupted when the protagonist took an action with negative consequences). On this region, the interaction was symmetrical, which is entirely in line with predictions based on the decision-theoretic channel acting alone. However, on the post-critical region the interaction was asymmetric. There was residual disruption when the protagonist had avoided an action with positive consequences (relative to a condition in which the protagonist avoided an action with negative consequences) but there were no residual disruption when the protagonist took an action with negative consequences (relative to a condition in which the protagonist took an action with positive consequences). This pattern of results suggests that readers quickly notice irrational actions when their attention is drawn to them, but the structural channel (which always anticipates action P and created a Theory of Mind blind spot in Experiment 1) helps readers to more easily recover from spotting a decision-theoretic anomaly.

Table 4
Experiment 2 mean reading times and regressions by analysis region and condition (Means averaged over subjects, standard errors in parentheses).

Utility of *q*/protagonist decision	First pass Regressions Out (%)	Regression Path (ms)	Regressions In (%)	Total Time (ms)
Critical region				
1. Desirable/P	18.8 (3.4)	928 (66)	13.4 (2.6)	880 (44)
2. Desirable/not-P	33.6 (3.9)	1301 (72)	24.1 (3.3)	1255 (53)
3. Undesirable/P	30.9 (3.8)	1150 (77)	18.8 (3.5)	1050 (49)
4. Undesirable/not-P	22.1 (3.5)	1059 (55)	13.2 (3.1)	1021 (45)
Post-critical region [a]				
1. Desirable/P	49.7 (5.9)	2565 (167)	n/a	1871 (79)
2. Desirable/not-P	71.1 (4.6)	3297 (217)	n/a	1944 (92)
3. Undesirable/P	59.2 (5.8)	2935 (190)	n/a	1898 (81)
4. Undesirable/not-P	50.8 (6.3)	2575 (165)	n/a	1907 (81)

[a] Regressions In were not applicable to post-critical region as this was always the final sentence of the vignette.

4. General discussion

People can track the beliefs and desires of other individuals, in order to rapidly predict their next actions. This fast component of Theory of Mind is usually very accurate in typical adult populations. In particular, people are very good at anticipating that other individuals will take actions that are subjectively beneficial. What we have demonstrated in this paper, though, is a blind spot in this fast and accurate Theory of Mind. According to our findings, people anticipate that a protagonist will take an action, regardless of its consequences. For example and quite strikingly, reading is perturbed when a protagonist avoids the oysters that would make her happy, but less perturbed when a protagonist orders the oysters that she knows will make her ill.

We predicted this asymmetry based on our suggestion that conditional sentences such as "if I have oysters for my main course, I will be very ill" build expectations through two parallel channels. The decision-theoretic channel (Bonnefon, 2009) builds an expectation that the protagonist will do (or not do) the action featured in the antecedent, when the consequences are desirable (or not desirable). The structural channel simply builds an expectation that the antecedent is true, irrespective of its consequences. When the protagonist takes an action that satisfies the expectations of at least one of the two channels, it is integrated into the ongoing discourse representation with little disruption. For example, disruptions are minimised when reading that the protagonist had oysters after stating "if I have oysters for my main course, I will be very ill", because the structural channel created the expectation that the protagonist would have oysters, despite the decision-theoretic expectation that she would not.

The existence of this structural channel is consistent with the findings of the ERP studies (Bonnefond & Van der Henst, 2009, 2013) we reviewed earlier in this article, but it is also consistent with most psychological theories of conditionals. Relevance theory (Sperber & Wilson, 1995) assumes that the most basic, initial mental representation of a conditional "if P then Q" is that there exist cases where P is true as well as Q (Sperber, Cara, & Girotto, 1995). Mental model theory (Johnson-Laird & Byrne, 2002) makes a similar assumption, and the suppositional theory of conditionals (Evans & Over, 2004) is based on the idea that the first step people take when assessing a conditional sentence, is to suppose that its antecedent is true. The existence of a structural channel building a quick expectation that P is true is also

Table 5
Experiment 2 ANOVA results for each region and relevant measure.

Region	Measure	Predictor	By participants			By items		
			$F_1(1,27)$	p	η_p^2	$F_2(1,23)$	p	η_p^2
Critical	Regression path	Utility	<1			<1		
		Decision	5.40	.028*	.167	4.3	.05*	.157
		Interaction	21.63	<.001*	.445	12.12	.002*	.345
	Total Time	Utility	<1			<1		
		Decision	24.50	<.001*	.476	9.22	.006*	.286
		Interaction	36.22	<.001*	.573	12.21	.002*	.347
	FPRO	Utility	<1			<1		
		Decision	<1			<1		
		Interaction	13.64	.001*	.336	9.64	.005*	.295
	Regressions In	Utility	<1			1.35	.257	.055
		Decision	1.14	.295	.040	<1		
		Interaction	11.17	.002*	.293	6.81	.016*	.228
Post-critical	Regression path	Utility	1.97	.171	.068	1.28	.269	.053
		Decision	2.45	.129	.083	4.02	.057	.149
		Interaction	21.58	<.001*	.444	8.51	.008*	.270
	Total Time	Utility	<1			<1		
		Decision	<1			1.22	.280	.050
		Interaction	<1			<1		
	FPRO	Utility	2.68	.113	.090	1.20	.285	.049
		Decision	4.79	.037*	.151	4.30	.050*	.157
		Interaction	15.21	.001*	.360	8.88	.007*	.279

* $p < .05$

consistent with the whole body of data suggesting that conditional sentences make the modus ponens inference (if P then Q, P, therefore Q) especially easy and quick to generate or verify (Barrouillet, Grosset, & Lecas, 2000; Evans, Newstead, & Byrne, 1993), to the point that it has been qualified as a cognitive reflex (Sperber, 2001). The co-existence of the structural and decision-theoretic channel also implies that when reading a conditional whose consequent is undesirable, people may have in mind both the true (P) and the false (not-P) antecedent of the conditional. Our data suggest that conditionals with desirable consequences result in readers constructing only an initial mental model representing 'P and Q' whereas conditionals with undesirable consequences may be 'fleshed out' to include the logically true possibilities 'P and Q' and 'not-P and Q' (Espino, Santamaria, & Byrne, 2009).

Critically, the coexistence of a decision-theoretic channel and a structural channel help explain why people can be mind-blind to avoidance desires. When reading a conditional rule such as "if action then undesirable consequences", people build both the expectation that the action will be taken (through the structural channel) and the expectation that it will not be taken (through the decision-theoretic channel). As a result, the Theory of Mind inference driven by the decision-theoretic channel is masked by the inference driven by the structural channel.

Clearly though, people do not routinely expect others to take detrimental actions, a fact that is well established in reasoning experiments (Bonnefon, Girotto, & Legrenzi, 2012; Bonnefon & Sloman, 2013). Given that people are blind to avoidance processes during online

Table 6
Paired comparisons between lexically identical regions of text following significant interaction effects in Experiment 2.

Region	Measure	Paired comparison	Mean difference	By participants			By items		
				t_1 (27)	p	d	t_2 (23)	p	d
Critical	Regression path	Desirable/P Undesirable/P	222 ms	3.28	.006*	.63	2.62	.030*	.54
		Desirable/not-p Undesirable/not-p	241 ms	3.11	.008*	.60	2.73	.024*	.56
	Total Time	Desirable/P Undesirable/P	170 ms	3.42	.004*	.65	2.27	.066	.48
		Desirable/not-p Undesirable/not-p	233 ms	4.03	<.001*	.77	3.34	.006*	.68
	FPRO	Desirable/P Undesirable/P	12.1 pp	3.03	.010*	.57	2.60	.032*	.53
		Desirable/not-p Undesirable/not-p	11.5 pp	2.41	.046*	.46	2.18	.080	.44
	Regressions In	Desirable/P Undesirable/P	5.4 pp	1.65	.222	.32	1.30	.410	.27
		Desirable/not-p Undesirable/not-p	10.9 pp	2.39	.048*	.47	2.81	.020*	.56
Post-critical	Regression path	Desirable/P Undesirable/P	370 ms	2.10	.090	.40	1.66	.222	.38
		Desirable/not-p Undesirable/not-p	722 ms	4.32	<.001*	.85	2.78	.022*	.68
	FPRO	Desirable/P Undesirable/P	9.5 pp	1.78	.140	.34	1.33	.394	.37
		Desirable/not-p Undesirable/not-p	20.3 pp	4.30	<.001*	.86	2.91	.016*	.88

* Bonferroni corrected p value < .05

processing, but well aware of these same avoidance desires in self-paced, offline reasoning tasks, it follows that the decision-theoretic channel must come to dominate at some point in time. On the other hand, even slow paced, offline reasoning tasks sometimes display the signature of the blind spot. For example, Bonnefon and Hilton (2004) found evidence that some participants considered that the antecedent of a conditional was true even when the consequent was neither desirable nor undesirable. Accordingly, it might prove difficult to pinpoint the exact time frame of the Theory of Mind blind spot to avoidance desires.

Our findings offer some insight into this difficulty. Whereas reading was always disrupted when protagonists took actions with undesirable consequences (both in the critical and post-critical region), Experiments 1 and 2 showed different patterns of disruption for avoidance desires. In Experiment 1, readers were initially unsurprised when protagonists avoided actions with desirable consequences, but showed some delayed signs of surprise in later regions of text. In contrast, when the materials drew attention to the surprising nature of this decision (Experiment 2), readers showed some initial surprise but quickly recovered, showing no residual signs of surprise in later regions of text. Accordingly, it appears that the structural channel can either delay the realization that the protagonist took an irrational action (Experiment 1), or neuter this realization when the attention is drawn to it (Experiment 2). Further research is thus necessary to attain a more precise knowledge of this time frame, either by focusing on online studies of intermediate stages of processing, or by using speeded versions of reasoning tasks performed with utility conditionals.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.actpsy.2015.05.009.

References

Apperly, I.A., Warren, F., Andrews, B.J., Grant, J., & Todd, S. (2011). Developmental continuity in theory of mind: Speed and accuracy of belief-desire reasoning in children and adults. *Child Development*, *82*, 1691–1703.

Barrouillet, P., Grosset, N., & Lecas, J. -F. (2000). Conditional reasoning by mental models: Chronometric and developmental evidence. *Cognition*, *75*, 237–266.

Bonnefon, J.F. (2009). A theory of utility conditionals: Paralogical reasoning from decision-theoretic leakage. *Psychological Review*, *116*, 888–907.

Bonnefon, J.F. (2012). Utility conditionals as consequential arguments: A random sampling experiment. *Thinking and Reasoning*, *18*, 379–393.

Bonnefon, J.F., Girotto, V., & Legrenzi, P. (2012). The psychology of reasoning about preferences and unconsequential decisions. *Synthese*, *187*, 27–41.

Bonnefon, J.F., & Hilton, D.J. (2004). Consequential conditionals: Invited and suppressed inferences from valued outcomes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 28–37.

Bonnefon, J.F., & Sloman, S.A. (2013). The causal structure of utility conditionals. *Cognitive Science*, *37*, 193–209.

Bonnefond, M., & Van der Henst, J.B. (2009). What's behind an inference? An EEG study with conditional arguments. *Neuropsychologia*, *47*, 3125–3133.

Bonnefond, M., & Van der Henst, J.B. (2013). Deduction electrified: ERPs elicited by the processing of words in conditional arguments. *Brain and Language*, *124*, 244–256.

Bonnefond, M., Van der Henst, J., Gougain, M., Robic, S., Olsen, M., Weiss, O., et al. (2012). How pragmatic interpretations arise from conditionals: Profiling the affirmation of the consequent argument with reaction time and EEG measures. *Journal of Memory and Language*, *67*, 468–485.

Espino, O., Santamaria, C., & Byrne, R.M.J. (2009). People think about what is true for conditionals, not what is false: Only true possibilities prime the comprehension of "if". *Quarterly Journal of Experimental Psychology*, *62*, 1072–1078.

Evans, J. St. B.T., Newstead, S.E., & Byrne, R.M.J. (1993). *Human reasoning: the psychology of deduction*. Hove: Erlbaum.

Evans, J. St. B. T., & Over, D. E. (2004) If. Oxford: Oxford University Press.

Ferguson, H.J., & Breheny, R. (2011). Eye movements reveal the time-course of anticipating behaviour based on complex, conflicting desires. *Cognition*, *119*, 179–196.

German, T.P., & Hehman, J.A. (2006). Representational and executive selection resources in 'theory of mind': Evidence from compromised belief–desire reasoning in old age. *Cognition*, *101*, 129–152.

Haigh, M., & Bonnefon, J.F. (2015). Eye movements reveal how readers infer intentions from the beliefs and desires of others. *Experimental Psychology*, http://dx.doi.org/10.1027/1618-3169/a000290.

Johnson-Laird, P.N., & Byrne, R.M.J. (2002). Conditionals: a theory of meaning, pragmatics, and inference. *Psychological Review*, *109*, 646–678.

Kovács, A., Téglás, M., & Endress, A.D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science*, *330*, 1830–1834.

Leslie, A.M., German, T.P., & Pollizi, P. (2005). Belief–desire reasoning as a process of selection. *Cognitive Psychology*, *50*, 45–85.

Malle, B.F., & Holbrook, J. (2012). Is there a hierarchy of social inferences? The likelihood and speed of inferring intentionality, mind, and personality. *Journal of Personality & Social Psychology*, *102*, 661–684.

Sperber, D. (2001). In defense of massive modularity. In E. Dupoux (Ed.), *Language, brain, and cognitive development* (pp. 47–57). Cambridge, MA: MIT Press.

Sperber, D., Cara, F., & Girotto, V. (1995). Relevance theory explains the selection task. *Cognition*, *57*, 31–95.

Sperber, D., & Wilson, D. (1995). *Relevance, communication and cognition*. Oxford: Blackwell (Original work published in 1986).