Applied Psycholinguistics (2024), **45**, 1219–1250
doi:10.1017/S0142716424000444

CAMBRIDGE UNIVERSITY PRESS

ORIGINAL ARTICLE

The cognitive underpinnings of irony comprehension: Fluid intelligence but not working memory modulates processing

Marianna Kyriacou and Franziska Köder

Department of Linguistics and Scandinavian Studies, University of Oslo, Oslo, Norway
Corresponding author: Franziska Köder; Email: franziska.koder@iln.uio.no

(Received 6 May 2024; revised 21 October 2024; accepted 7 November 2024)

Abstract

The comprehension of irony involves a sophisticated inferential process requiring language users to go beyond the literal meaning of an utterance. Because of its complex nature, we hypothesized that working memory (WM) and fluid intelligence, the two main components of executive attention, would be involved in the understanding of irony: the former by maintaining focus and relevant information active during processing, the latter by disengaging irrelevant information and offering better problem-solving skills. In this eye-tracking reading experiment, we investigated how adults ($N = 57$) process verbal irony, based on their executive attention skills. The results indicated a null (or indirect) effect for WM, while fluid intelligence directly modulated the comprehension and processing of irony during reading. As fluid intelligence is an important individual-difference variable, the findings pave the way for future research on developmental and clinical populations who tend to struggle with nonliteral language.

Keywords: Executive attention; fluid intelligence; irony; nonliteral language processing; working memory

Introduction

Understanding an utterance such as "interesting movie!" when the speaker is visibly disengaged and uninterested requires bridging the gap between what the speaker literally says (i.e., the movie is interesting) and what they intend to communicate to their conversation partner (i.e., disappointment that their expectation of watching an interesting movie was not met). Interpreting nonliteral language, particularly verbal irony, can be challenging because cues to the intended meaning can be subtle and the literal meaning can be contradictory and require suppression. Previous research has shown that irony is one of the most complex pragmatic phenomena. Children younger than six years of age (Falkum & Köder, 2020), as well as clinical

© The Author(s), 2025. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (https://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided that no alterations are made and the original article is properly cited. The written permission of Cambridge University Press must be obtained prior to any commercial use and/or adaptation of the article.

https://doi.org/10.1017/S0142716424000444 Published online by Cambridge University Press

populations such as individuals with autism spectrum disorder (e.g., Caillies et al., 2014; Martin & McDonald, 2005; Wang et al., 2006), are known to struggle with ironic meanings.

A classic explanation for why irony may be harder to process than literal language is that the listener needs to first detect that a literal interpretation is incongruous with the context and subsequently infer the ironic interpretation the speaker intends to convey (Grice, 1975). Other theories conceptualize irony as a form of pretense or echo (Clark & Gerrig, 1984; Wilson & Sperber, 2012). To grasp the meaning of "interesting movie!", for instance, would require the listener to recognize that the speaker is "echoing" or "pretending to entertain" the belief that the movie would be great, while simultaneously expressing a dismissive attitude toward that (from their current perspective) ludicrously false belief.

While ensuing theoretical accounts have stressed that various phrase-, context-, and reader-related factors can speed up activation of ironic relative to literal meanings (e.g., Gibbs, 1994; Giora, 2003; Katz & Ferretti, 2001; Pexman, 2008; Sperber & Wilson, 1981), the role of individual differences in irony processing requires more research (see review by Kałowski et al., 2023). Evidence suggests that theory of mind and mentalizing skills (e.g., Filippova & Astington, 2008; Nilsen et al., 2011), i.e., cognitive abilities related to understanding and attributing mental states, as well as emotion understanding (e.g., Jacob et al., 2016; Nicholson et al., 2013), positively affect irony comprehension. Other sources of individual differences related to executive functions or executive attention abilities (i.e., domain-general cognitive processes essential for managing and regulating thoughts and behavior), also seem to play a role in irony processing, yet evidence to date is limited. Research on children's irony development suggests that working memory (WM), inhibitory control, and inferential reasoning abilities, as well as overall cognitive abilities, facilitate irony comprehension (Caillies et al., 2014; Filippova & Astington, 2008; Godbee & Porter, 2013). However, it remains unclear whether executive attention abilities are also particularly taxed in the processing of irony by adults. Additionally, there is conflicting evidence regarding the role of WM in this context (Olkoniemi & Kaakinen, 2021), while the role of fluid intelligence (Gf)—another key component of executive attention—has been largely overlooked.

To gather new, more fine-grained insights into the basic cognitive mechanisms involved in the processing of irony, we designed an eye-tracking reading experiment exploring the roles of working memory as well as fluid intelligence in neurotypical adults.

The comprehension of irony

While irony is sometimes defined broadly as including jocularity, sarcasm, hyperbole, rhetorical questions, and understatements (e.g., Gibbs, 2000; Recchia et al., 2010), we use a narrower definition of irony as it is unclear whether these different communicative phenomena all rely on the same cognitive mechanisms. We focus on instances of irony where a speaker says something literally positive such as "this is great" or "interesting movie" to express a critical, mocking, or contemptuous attitude toward that thought (Clark & Gerrig, 1984; Wilson & Sperber, 2012).

https://doi.org/10.1017/S0142716424000444 Published online by Cambridge University Press

Multiple studies and a variety of research methodologies demonstrate that ironic utterances lead to differential processing compared to literal utterances. Event-related potential studies show greater N400 or P600 amplitudes for ironic phrases, two components linked with the identification of inconsistencies between linguistic utterances and the surrounding context in pragmatics, thus suggesting more effortful meaning integration as compared to literal phrases (Cornejol et al., 2007; Filik et al., 2014; Regel et al., 2010, 2011; Spotorno et al., 2013). Findings from fMRI studies show that irony causes greater brain activity, with respect to both locality and magnitude of activation (Akimoto et al., 2014; Bosco et al., 2017; Obert et al., 2016; Shibata et al., 2010; Spotorno et al., 2012), while complementary findings from reading studies demonstrate that a phrase such as "this is useful" is read significantly more slowly when intended in an ironic versus a literal sense (Dews & Winner, 1999; Giora et al., 1998). Further data from eye-tracking show that this irony reading cost emerges in late measures, such as total reading time (Au-Yeung et al., 2015; Filik et al., 2014, 2018; Filik & Moxey, 2010), and regression likelihood, as readers are more likely to go back to reread ironic as opposed to literal phrases (Kaakinen et al., 2014; Olkoniemi et al., 2016; Olkoniemi, Johander, et al., 2019). Reading studies incorporating explicit comprehension questions further report reduced accuracy for ironic compared to literal utterances (e.g., Kaakinen et al., 2014; Olkoniemi et al., 2016; Olkoniemi, Johander, et al., 2019; Olkoniemi, Strömberg, et al., 2019). As late eye-tracking reading measures are believed to capture meaning integration processes and reanalysis (Conklin et al., 2018), the collective findings suggest that irony is resolved in later stages of processing, often after laborious reanalysis (i.e., regressive rereading), with processing costs being evident in both implicit (reading behavior) and explicit (overt comprehension) measures.

Notably, the irony processing cost can be mitigated by certain factors. For instance, frequently encountering a phrase in an ironic sense can lead to a default ironic interpretation of said phrase, thus facilitating its comprehension (Filik et al., 2014; Giora, 2003; Giora et al., 2015). Context and cultural expectations can further modulate processing effort and speed, making irony considerably faster to read, and at times even faster than literal phrases (Katz et al., 2004; Ronderos et al., 2023; Spotorno & Noveck, 2014). This suggests that (at least part) of the irony processing cost observed in aforementioned studies can be offset by stronger phrasal and contextual cues. Importantly, beyond these influences, individual differences stemming from variability in executive attention abilities could explain some of the observed differences in the processing and comprehension of irony.

Executive attention and irony processing

Executive attention (alternatively referred to as attention control, or under different frameworks as executive function) is paramount to most models of higher-order cognition (e.g., Atkinson & Shiffrin, 1968; Baddeley, 1998; Draheim et al., 2021; Posner & DiGirolamo, 1998; Shipstead et al., 2016). It refers to the ability to control one's thoughts and behavior for task performance purposes, such as resisting distraction from unrelated external events (e.g., disruptive noises during work), or unrelated internal thoughts (e.g., thinking about lunch during a morning meeting).

https://doi.org/10.1017/S0142716424000444 Published online by Cambridge University Press

Traditionally, the more deconstructive approach to executive function assumes that mechanisms such as WM, inhibition, and shifting tap into domain-specific aspects of attention control (e.g., Friedman & Miyake, 2017; Miyake et al., 2000). However, the validity of these constructs has come under scrutiny, as many studies under this framework have failed to find a (strong) association between different tasks designed to measure the same constructs, casting doubts on the psychometric validity of both the constructs themselves and the tasks used to measure them. For instance, low intercorrelations have been reported between distinct inhibition tasks (Paap et al., 2015; Rey-Mermet et al., 2018; Rouder & Haaf, 2019), implying that inhibition operations employed in one task are different from those employed in another.

An alternative approach has emerged, which posits that executive attention is a domain-general construct. Under this approach, attention control is (at least partly, if not fully) mediated by two main functions: working memory and fluid intelligence. Working memory is assumed to be responsible for maintaining relevant information active during processing (as well as maintaining attention to the task), while fluid intelligence is responsible for information disengagement once this has been proven to be irrelevant or incorrect for the problem at hand (Burgoyne & Engle, 2020; Engle, 2002, 2018; Shipstead et al., 2016). The cognitive tasks developed or adapted under this framework, by Engle and collaborators, have better psychometric properties and reliability compared to previous tasks developed under the more traditional executive function framework (Draheim et al., 2021, 2023). This is because these tasks eliminate, among other things, confounds from speed-accuracy trade-offs, and correlate to a greater extent with both WM and fluid intelligence, thus lining up with a more unitary, general-purpose attention control system (Draheim et al., 2019, 2021).

For our study, we adopt this domain-general view of executive attention to investigate its potential involvement in the processing of irony. When phrases are not markedly ironic (Giora, 2003; Giora et al., 2015), processing irony can demand significant attentional resources, as it requires overcoming a literal-compositional interpretation by taking into account contextual factors (Gibbs, 1994; Katz et al., 2004) and speaker attitudes (Utsumi, 2000). We therefore hypothesized that both irony comprehension and processing may depend on individual attentional capacities. Below, we discuss the constructs of WM and fluid intelligence in more detail, exploring their potential impact on the comprehension of ironic utterances.

Working memory

Working memory is a limited-capacity system where memory and attention interact to facilitate the simultaneous processing and storage of information (Baddeley & Hitch, 1974). Greater WM capacity comes with processing advantages as it can simultaneously accommodate larger pieces of information, as well as improve one's ability to maintain attention to the task by ignoring both internal and external distractors (Engle, 2018). For successful reading comprehension, WM is important for several reasons. A reader needs to direct and maintain attention to the text, allocate attentional resources to the mechanics of reading (oculomotor control, decoding, word recognition, etc.), as well as to higher-level processes such as

https://doi.org/10.1017/S0142716424000444 Published online by Cambridge University Press

retaining pertinent textual information active in memory to form mental representations of the content (Martin et al., 2020; Shin, 2020). These representations are then subject to constant updating from further incoming input, as the reader gathers new pieces of information from subsequent parts of the text and engages in an ongoing integration process (Burgoyne et al., 2022). Indeed, a strong correlation between WM and reading comprehension has consistently shown that better WM predicts better comprehension (Butterfuss & Kendeou, 2018; Carretti et al., 2009; Daneman & Carpenter, 1980; Turner & Engle, 1989).

Unlike in general reading comprehension, the role of WM in the processing of verbal irony (during reading) has only recently begun to attract attention. It could be argued that WM may facilitate the processing of written irony, as the reader would need additional attentional resources to register a discrepancy between the context and the literal meaning of an ironic utterance, requiring additional meaning disambiguation and selection processes. Some eye-tracking studies have shown that readers with greater WM capacity read ironic phrases significantly more *slowly* in early measures, and specifically during first-pass reading (Kaakinen et al., 2014; Olkoniemi et al., 2016), while readers with reduced WM go back (regress) to reread ironic phrases after they have read later parts of the text (Olkoniemi et al., 2016; Olkoniemi, Johander, et al., 2019). These findings would suggest that greater WM capacity leads to faster irony detection during reading, although notably this is *not* accompanied by faster reading times. Instead, earlier detection seems to initiate ambiguity resolution at an earlier timepoint. Lower WM capacity, on the other hand, may yield compensatory strategies and reanalyzes, as indicated by the tendency to reread ironic phrases.

However, the role of WM in irony processing is not firmly established, as some studies failed to find an effect altogether (Olkoniemi, Strömberg, et al., 2019; Parola & Bosco, 2022). Olkoniemi, Strömberg et al. (2019) considered whether low task demands arising from stimuli with short contexts might explain the absence of an observed effect, although Parola and Bosco (2022) failed to find an effect despite using longer contexts.

These findings, therefore, cast some doubts as to whether WM is always particularly taxed during irony comprehension, over and above what is required for the processing of literal utterances. The degree of its involvement may, for instance, depend on utterance "defaultness" (Giora et al., 2015), with markedly ironic phrases requiring fewer WM resources relative to more ambiguous, context-dependent phrases. Overall, the conflicting evidence creates a need for further investigation into the role of WM in irony processing.

Fluid intelligence

Fluid intelligence, the other component of executive attention, relates to one's ability to solve novel problems and reason with novel information (Cattell, 1943; Horn & Cattell, 1966), where being able to disengage irrelevant information, such as disproven hypotheses or incorrect solutions, can free up attentional resources (Engle, 2018). These resources can then be redirected toward assessing viable options, thus maximizing both task performance and chances of success. With reference to reading, fluid intelligence may help the disengagement of

https://doi.org/10.1017/S0142716424000444 Published online by Cambridge University Press

activated, but contextually inappropriate meanings (e.g., the nontarget meaning(s) of a polysemous word), or with the disengagement of textual representations that are invalidated in later parts of the text. Although underexplored as an aspect of reading comprehension, recent findings showcase that fluid intelligence strongly correlates with reading comprehension, while WM may not have a *direct* influence beyond its association with fluid intelligence (Martin et al., 2020). That is, the correlation between WM and reading comprehension was found to be accounted for by variance that WM shared with fluid intelligence.

Indeed, under this framework of executive attention, WM and fluid intelligence are believed to be correlated (yet distinct) constructs (Kane et al., 2005; Oberauer et al., 2005), as they both rely on the same ability to control attention: WM to maintain and fluid intelligence to disengage information. For instance, when solving a problem, one would need to keep a working hypothesis active in memory (e.g., Engle, 2018), while hypothesis testing itself also requires keeping track of a prediction (i.e., probable solution) to be compared against available evidence (Mashburn et al., 2023). Similarly, in complex WM tasks that include not only a memory but also an unrelated processing component (e.g., a judgment task designed to tax the processing system), elements of fluid intelligence, particularly the ability to disengage, are crucial. For example, one would need to quickly disengage resources from the judgment task, in order to focus and perform well on the recall (memory) component. Nevertheless, as WM and fluid intelligence are still considered distinct mechanisms, if fluid intelligence exerts independent influence on the processing of irony, as it appears to be the case in reading comprehension (Martin et al., 2020), its effects should be visible irrespective of WM influences.

Interestingly, some preliminary findings have already linked fluid intelligence with the ability to detect ironic praise (i.e., negative statements expressing positive attitudes) in aptitude tests (Bruntsch & Ruch, 2017), but it remains to be seen whether and to what extent fluid intelligence influences the on-line processing of irony, and particularly of ironic criticism (i.e., positive statements expressing negative attitudes), which is the most common form of ironic language use (Kreuz & Link, 2002). For our purposes, it could be argued that fluid intelligence may facilitate the processing of irony because of overall better problem-solving abilities, or because of an enhanced ability to disengage nontarget meanings, which would make the disengagement of the literal meaning of ironic utterances more efficient. Furthermore, assuming that problem-solving operations would be mobilized once a "problem" has been identified, we would expect fluid intelligence to influence processing only *after* irony has been detected, or at least as soon as the literal meaning is understood to be inappropriate in the context and in need of reconsideration. Therefore, unlike WM whose effects might appear early during processing, fluid intelligence might affect later processing stages that tap into more comprehensive disambiguation and meaning integration processes. Any effects should therefore emerge in measures like total reading time and regression likelihood, which incorporate rereading and reanalysis as part of meaning integration processes, or in comprehension questions that explicitly tap into the interpretation of irony. However, as we have argued with respect to WM, in cases of more default ironic meanings, or where context is strong enough to directly activate an ironic meaning, fluid intelligence might not play a significant role, as problem-

https://doi.org/10.1017/S0142716424000444 Published online by Cambridge University Press

solving skills would not be required as much (i.e., no alternative hypotheses need to be considered or discarded).

The present study

From the discussion above, it appears that both components of executive attention might play a role in the processing of irony, especially when utterances do not render an ironic reading by default but are instead subject to contextual inferences. In this study, we examine the extent to which WM and fluid intelligence may influence the on-line reading and comprehension of verbal irony. To investigate this, we designed an eye-tracking while reading task that involved reading stories containing (the same) target phrases intended ironically or literally. We considered both early and late eye-tracking reading measures (e.g., first-pass reading time, total reading time, regression probability), as well as accuracy and response time to explicit comprehension questions. Working memory capacity and fluid intelligence were used as predictors in analyses. Supplementary Materials including stimuli, data, analysis code, and model outputs are available here: https://osf.io/4f7xm/.

If WM is indeed involved in the processing of irony (an effect that cannot be taken for granted based on previous findings), higher WM capacity may speed up the detection of irony, potentially resulting in longer first-pass reading times as readers contemplate an alternative meaning earlier during processing. Higher fluid intelligence, on the other hand, may affect irony processing in later stages of processing after irony has been detected, indicating an involvement in meaning disambiguation (i.e., activation and selection of ironic meaning and disengagement of competing nontarget ones) and meaning integration through the engagement of general problem-solving skills. Therefore, higher fluid intelligence may be evident in late measures such as total reading time and regression probability, although it is difficult to predict the direction of the effect (e.g., faster reading times/fewer regressions vs. longer reading times/more regressions) as successful problem-solving may be susceptible to speed-accuracy trade-offs: i.e., spending more processing time for more accurate problem-solving. On the other hand, it is also possible that greater fluid intelligence may speed up processing because of increased problem-solving efficiency relative to lower fluid intelligence.

The aims of this study were threefold. First, we sought to establish a role of WM in the processing of irony, and second, to examine if fluid intelligence is also an important predictor. Third, we explored whether these constructs influence processing independently at different stages, reflecting distinct involvement in meaning activation, and ultimately in meaning selection and integration in the context.

Methods

Participants

Sixty native speakers of Norwegian took part in the main experiment, conducted at the Socio-Cognitive Laboratory at the University of Oslo. Three of them were removed from analyses due to camera calibration problems, leaving 57 participants

https://doi.org/10.1017/S0142716424000444 Published online by Cambridge University Press

in total (mean age = 23.19, *SD* = 3.88, females = 38; males = 19). All participants were Norwegian native speakers, between 18 and 35 years of age, and without known cognitive or language impairments. Participants provided informed consent and received compensation in the form of a gift card for their participation. Ethical approval for the study was granted by the Norwegian Agency for Shared Services in Education and Research (SIKT; reference number 478374; project preregistration: https://osf.io/xhd7g).

Pilot study and power analysis

Sample size was determined based on power analyses conducted on pilot data. For the pilot study, we recruited eleven participants who fulfilled the inclusion criteria. Power analyses were carried out via model simulations on a comprehensive variety of different measures: first pass and total reading time of target phrase region, regression probability to context region, and response accuracy to inference questions on irony. Model structures were specified as in the main analyses and simulations were computed using *mixedpower* (Kumle et al., 2018) in R, version 4.2.2 (R Core Team, 2023). Simulations were compared using three different sample sizes (30, 60, and 100 participants), with the critical *t*/*z* value set at 2, and single simulations set to 1000 (Kumle et al., 2021). The outcome of the power analyses confirmed that 60 participants provided enough power ($\geq$0.8) to reliably detect an effect if existent, with no notable differences between the 60- and 100-participant groups. It is worth mentioning, however, that effects of the power analysis were not always on par with those of the main analyses, likely due to the much smaller number of observations. This sample size is similar to those of previous similar studies (e.g., Kaakinen et al., 2014; Olkoniemi et al., 2016; Olkoniemi, Johander, et al., 2019).

Materials

The eye-tracking reading task included twenty-four pairs of stories (N = 48) written in Norwegian. The stories were on average 721.24 characters long (*SD* = 47.56) and always described a situation involving one or two fictional characters. At some point, one character would utter a phrase (referred to as target phrase) that would be intended ironically in the irony condition (N = 24), and literally in the literal condition (N = 24). Each story was split into three Regions of Interest: (a) the *context region,* which included all text leading up to (but not including) the target phrases, (b) the *target phrase region*, and (c) the *spillover region* (all the content directly following the target phrases and up to the end of the story). The target phrase and spillover regions were identical across conditions, whereas the context region contained (minimal) phraseological differences, which were necessary to convey the intended meaning of the target phrases (i.e., literal vs. ironic). For example stimuli translated from Norwegian to English, see Table 1. Twelve filler stories with a similar structure, some containing untruthful statements or statements that required complex inferences, were included to minimize irony's markedness.

https://doi.org/10.1017/S0142716424000444 Published online by Cambridge University Press

All story pairs were accompanied by (the same) two YES/NO comprehension questions. Question 1 tapped into general content comprehension, while Question 2 tapped into the interpretation of the target phrases. The second required drawing inferences about the implicit feelings/thoughts/intentions of the speaker who uttered the target phrase in each story (see example inference question in Table 1), thus, the correct answer would depend on the condition. Question 2 will be henceforth referred to as inference question. Inference questions were phrased in such a way as to elicit a good balance of negative and positive answers in the irony versus the literal condition. Filler stories were also accompanied by two questions of the same kind.

To ensure the stories successfully warranted the intended meaning of the target phrases in each condition, we normed them via two counterbalanced rating questionnaires, where only one version from each pair appeared in either questionnaire. Thirty-one participants who did not take part in the eye-tracking study (mean age = 34.87, $SD = 15.61$, females = 19, males = 12) rated how ironic versus literal the target phrases were in their respective story contexts on a Likert scale from 1 (literal) to 7 (ironic). The same participants rated the stories for naturalness, also on a Likert scale from 1 (very unnatural) to 7 (very natural). The norming data confirmed that target phrases in the irony condition were rated as significantly more ironic ($M = 6.41$, $SD = 0.57$) than (the same) target phrases in the literal condition ($M = 1.96$, $SD = 0.91$; $t(48) = 20.18$, $p < 0.001$), but all stories were rated to be equally natural ($M_{\text{Ironic}} = 5.33$, $SD = 0.36$; $M_{\text{Literal}} = 5.35$, $SD = 0.52$; $t(48) = 0.14$, $p = 0.89$). Another 17 participants (mean age = 33.88, $SD = 15.27$; females = 14) rated the target phrases in isolation (i.e., without their story contexts) on a Likert scale from 1 (literal) to 7 (ironic). The results showed that without context the phrases were considered literal ($M = 2.24$, $SD = 0.89$), thus excluding a default ironic interpretation.

Procedure

The whole experimental procedure lasted approximately 60 minutes and consisted of an eye-tracking reading task (ca. 40 minutes) and two cognitive tasks (ca. 20 minutes in total) tapping into WM capacity and fluid intelligence, respectively. Participants always started with the reading task, while the order of the administration of the cognitive tasks was counterbalanced across participants to account for fatigue effects.

Eye-tracking reading task

To track participants' eye-movements while reading, we used an SR Research EyeLink 1000+ desktop-mounted eye-tracker, with a sampling rate of 1000 Hz (SR Research, Ontario, Canada). Participants were seated in front of a computer monitor and a chin- and forehead-rest was used to minimize head movement and ensure better calibration. Calibration was performed using a nine-point grid, and recalibration was performed after breaks and whenever necessary.

The stories were divided into two counterbalanced lists using a Latin square design, so that each participant read a given target phrase only once, in either the

https://doi.org/10.1017/S0142716424000444 Published online by Cambridge University Press

https://doi.org/10.1017/S0142716424000444 Published online by Cambridge University Press

Table 1. Example stimuli stories and inference questions across Phrase Type (Ironic and Literal), translated to English

	Phrase Type	
	Ironic	Literal
Context region	Sarah and Jack have different approaches to life. Jack is a stickler for punctuality, he is always super organized, and well-prepared in advance; something that Sarah finds quite irritating at times. When they fly together, Jack always insists that they get to the airport at least 4 hours ahead of their departure to allow for any unexpected delays. Yesterday, they were flying out to the Fiji Islands for holidays. When they got to the airport, they found out that their flight had been **postponed for 3 hours** due to scheduling changes, although they hadn't been informed by the airlines.	Sarah and Jack have different approaches to life. Jack is a stickler for punctuality, he is always super organized, and well-prepared in advance; something that Sarah finds quite irritating at times. When they fly together, Jack always insists that they get to the airport at least 4 hours ahead of their departure to allow for any unexpected delays. Yesterday, they were flying out to the Fiji Islands for holidays. When they got to the airport, they found out that their flight had been **brought forward by 3 hours** due to scheduling changes, although they hadn't been informed by the airlines.
Target phrase region	**"Well, good thing we got here early!",**	**"Well, good thing we got here early!",**
Spillover region	commented Sarah. Jack mumbled something along the lines of, "They ought to have told us, really", and went on to check his emails for any missed correspondence.	commented Sarah. Jack mumbled something along the lines of, "They ought to have told us, really", and went on to check his emails for any missed correspondence.
Inference question	Was Sarah glad that she and Jack had arrived at the airport earlier?	
Correct answer	No	Yes

Note: Biasing contextual information in the context region, and the target phrases are presented in bold in the examples above but were in no way demarcated for participants.

irony (12 stories) or the literal condition (12 stories). The same 12 filler stories were used in both lists. Target phrases were never placed at the end of a sentence/line to avoid influence from wrap-up or saccadic programming operations. Participants were instructed to read the stories as fast as possible but for comprehension, to avoid unnecessary rereading wherever possible, and to press ENTER after reading each story. At that point, the story disappeared from the screen and participants were (sequentially) presented with the two comprehension questions. To answer the questions, participants were instructed to use the corresponding keys for YES/NO on the keyboard. Once the comprehension questions were answered, there was a drift correction before the next story appeared on the screen. The stories were presented in full on "one screen" at a time using triple line spacing, in black font (Courier New, size 16) over a white background. The order of the stories was randomized per participant.

Working memory task

To measure participants' WM capacity, we used Foster et al.'s (2015) shortened version of the Symmetry Span task (henceforth referred to as SSPAN). The task was translated into Norwegian at our lab. In this task, participants alternate between performing an irrelevant distractor task (i.e., judgment task) and a target memory task. The distractor task involves judging whether a geometrical pattern is symmetrical along the vertical axis, while the memory task involves retaining in memory and recalling which cell within a blank grid turned red. The distractor task alternates with the block presentation for a varying number of alternations, with a single block turning red within the grid upon each presentation. At the end of this alternating sequence, participants view a blank grid and must click on the locations of the red blocks in the order they had appeared within the sequence. The number of the to-be-remembered red cells varied by trial, ranging from two to five. The task consisted of three blocks, each block yielding a maximum of 14 points (42 in total). The task is scored for the total number of items correctly recalled. This is a complex span task as it involves a distractor (processing component) as well as a recall (memory) component. We chose this type of WM task because it is a nonverbal task and has been found to modulate the processing of sarcasm (a subtype of irony) in previous work (Olkoniemi, Johander, et al., 2019).

Fluid intelligence task

To measure participants' fluid intelligence, we used a computerized version of Raven's Advanced Progressive Matrices (APM III; Pearson, 2015). This is a nonverbal task containing 23 items/problems. Each item consists of a black-and-white geometric pattern with a piece missing. Participants need to identify which one out of the 8 options provided below correctly completes the pattern. Participants were given 10 minutes to solve as many problems as they could. The task was scored for the total number of correct answers.

https://doi.org/10.1017/S0142716424000444 Published online by Cambridge University Press

Table 2. Participants' scores on the SSPAN (working memory) and Raven's (fluid intelligence) tasks

	SSPAN	Raven's
Mean score	29.15 [/42]	33.33 [/100]
SD	8.05	15.15
Range	7–41	4.34–65.21

Note. SSPAN scores reflect raw partial scores while Raven's scores are presented in percentages. Both variables were centered before analyses.

Results

All data analyses were carried out in R version 4.3.1 (R Core Team, 2023). As an index of WM capacity, we calculated SSPAN task partial scores (where score credit is also given to partially correct items, for instance, when a participant could recall the location and sequence of some but not all of the cells that turned red in the grid; Conway et al., 2005), using the *englelab* package (Tsukahara, 2022). As an index of fluid intelligence, we calculated Raven's scores as percentage of correct answers. The descriptive statistics for both cognitive tasks are provided in Table 2.

For the reading task, we examined both early and late eye-tracking measures. Specifically, for the target phrase and spillover regions, we examined *first-pass reading time* (duration of all fixations during first pass) and *first-pass gaze duration* (duration of all fixations and refixations during first pass), both early measures, as well as late measures: i.e., *go-past reading time* (duration of all fixations and refixations in the region of interest including time spent revisiting previous regions), and *total reading time* (duration of all fixations and refixations). Additionally, for the target phrase region and the context region, we examined *regression probability* (i.e., how likely it was for these regions to be revisited from subsequent parts of the text), also a late measure. We did not analyze reading time measures for the context region as effects would be influenced by lexico-syntactic differences (e.g., differences in word frequency, word length, and polarity), given that this region varied slightly across Phrase Type. Regressions reflect late processing stages typically associated with reanalysis or meaning integration (global comprehension) and as such they are less susceptible to influences from lower-level lexico-syntactic variations. For the comprehension task, we examined both *accuracy* and *response time* to the inference questions (i.e., Question 2), as these tapped into the interpretation of the target phrases. Response accuracy to general content questions (i.e., Question 1) on both filler and experimental items was high (81% and 87%, respectively), demonstrating that participants were attentive to the reading task. Eye-tracking data loss due to camera tracking issues or the removal of outliers ranged between 0.44 and 3.00% across various eye-tracking measures and regions of interest.

The data were analyzed using mixed-effects models and the *lme4* package, version 1.1-34 (Bates et al., 2014). As fixed effects in all models, we specified Phrase Type (a two-level factor: Irony vs. Literal), Raven's scores and SSPAN scores, both continuous variables and henceforth referred to as fluid intelligence and WM scores, respectively, and Trial Index (i.e., trial presentation number, also a continuous variable). Trial Index was included as a control variable to account for learning/fatigue

https://doi.org/10.1017/S0142716424000444 Published online by Cambridge University Press

effects as the experiment progressed. We further specified an interaction between Phrase Type and both fluid intelligence and WM scores to check whether these predictors were particularly implicated in the processing of irony. Phrase type was effect-coded (using *contr.sum()*in R) with the literal condition being set as the reference level. Fluid intelligence and WM scores, as well as Trial Index, were centered and scaled. There was no indication of collinearity between fluid intelligence and WM scores ($r = 0.28$ and $\kappa = 1.33$) and the variance inflation factor values of all fixed effects in the models reported were low (i.e., below 5). We therefore included both scores as predictors in the models.

Maximal models included by-participant and by-item correlated random intercepts and slopes for Phrase Type and Trial Index. For items, we further specified random intercepts of fluid intelligence and WM scores in interaction with Phrase Type and Trial Index. We selected final models with a simplified random effect structure that did not lead to convergence issues or singular fits, first by removing interactions between random effects, and subsequently by removing random effects whose variance estimates were 0. Below we report only significant effects, but the complete outputs of the final models are provided in the appendix (Table A1). As we were primarily interested in potential effects of fluid intelligence and WM scores, models in which either of these predictors was significant were compared to null models lacking the respective predictor as a fixed effect. Full and null model comparisons were performed using *anova()* in R.

All reading and reaction time measures were log-transformed and analyzed with linear mixed-effects models to normalize distribution. Binary measures (i.e., regression probability and accuracy on inference questions) were calculated using logistic regression (Jaeger, 2008). Means across Phrase Types are provided in Table 3.

Eye-tracking reading measures

In the target phrase region, Phrase Type was significant in the first-pass gaze duration ($\beta = 0.02$, $SE = 0.01$, $t = 1.98$, $p = 0.047$), go-past reading time ($\beta = 0.03$, $SE = 0.01$, $t = 2.56$, $p = 0.01$), total reading time ($\beta = 0.04$, $SE = 0.01$, $t = 2.75$, $p = 0.006$), and regression probability ($\beta = 0.22$, $SE = 0.08$, $z = 2.72$, $p = 0.006$), showing that reading times were significantly slower and regression probability significantly higher for ironic as opposed to literal phrases. In the spillover region, Phrase Type was only significant in the first-pass reading time ($\beta = -0.10$, $SE = 0.02$, $t = -3.67$, $p < 0.001$), whereby reading times were *faster* in the irony as opposed to the literal condition. This likely reflects a processing trade-off, as ironic phrases elicited more and longer fixations in the target phrase region compared to literal phrases. In the context region, Phrase Type was not significant in regression probability ($p = 0.07$). Trial index was significant in most models (see appendix, Table A1 for coefficients), indicating that reading times or regression probability reduced as experience with the experiment increased.

Comprehension task measures

A coding error in building the experiment resulted in having to exclude half the responses from the literal (but not the irony) condition in the accuracy model[1].

https://doi.org/10.1017/S0142716424000444 Published online by Cambridge University Press

Table 3. Mean reading times and regression probability across regions of interest and explicit comprehension measures, collapsed across Phrase Type

	Phrase Type			
	Literal		Ironic	
	Mean	*SE*	Mean	*SE*
Target phrase region				
First pass reading time	791.00	17.73	819.00	18.09
First pass gaze duration	**939.00**	21.41	**969.00**	20.20
Go-past reading time	**1002.00**	24.41	**1070.00**	25.44
Total reading time	**1051.00**	23.54	**1151.00**	25.74
Regression probability	**0.19**	0.01	**0.25**	0.01
Spillover region				
First pass reading time	**3918.00**	86.51	**3580.00**	89.86
First pass gaze duration	4823.00	85.21	4846.00	83.42
Total reading time	4823.00	85.21	4846.00	83.42
Context region				
Regression probability	0.54	0.02	0.61	0.02
Comprehension task				
Accuracy on inference questions	**0.89**	0.02	**0.74**	0.02
Response time to inference questions	**3407.00**	94.26	**3965.00**	118.89

Note. Reading and Response time means are reported in milliseconds, while Accuracy and Regression probability as probability 0–1. Values in bold denote a significant difference across Phrase Type based on model outputs.

This model was then simplified to allow convergence by removing interactions between Phrase Type and fluid intelligence as well as WM scores. Phrase Type was significant in both the accuracy ($\beta = -0.57, SE = 0.17, z = -3.23, p = 0.001$) and response time to inference questions ($\beta = 0.05, SE = 0.01, t = 3.19, p = 0.001$), with accuracy being significantly lower and response time significantly slower in the irony as opposed to the literal condition. Again, Trial Index was significant, leading to increased accuracy and faster response time over the course of the experiment. Of note, Phrase Type was not significant in response accuracy on the general story comprehension question ($\beta = 0.20$, $SE = 0.31$, $z = 0.65$, $p = 0.51$) and comprehension scores were (equally) high in both conditions ($M_{LITERAL} = 0.87$, $M_{IRONY} = 0.87$).

Fluid intelligence effects

Fluid intelligence was significant as a main effect in the accuracy ($\beta = 0.03$, $SE = 0.01$, $z = 3.07$, $p = 0.002$) and response time ($\beta = -0.005$, $SE = 0.002$, $t = -2.51$, $p = 0.01$) to inference questions: as fluid intelligence increased, accuracy increased and response time became faster. There was a significant

https://doi.org/10.1017/S0142716424000444 Published online by Cambridge University Press

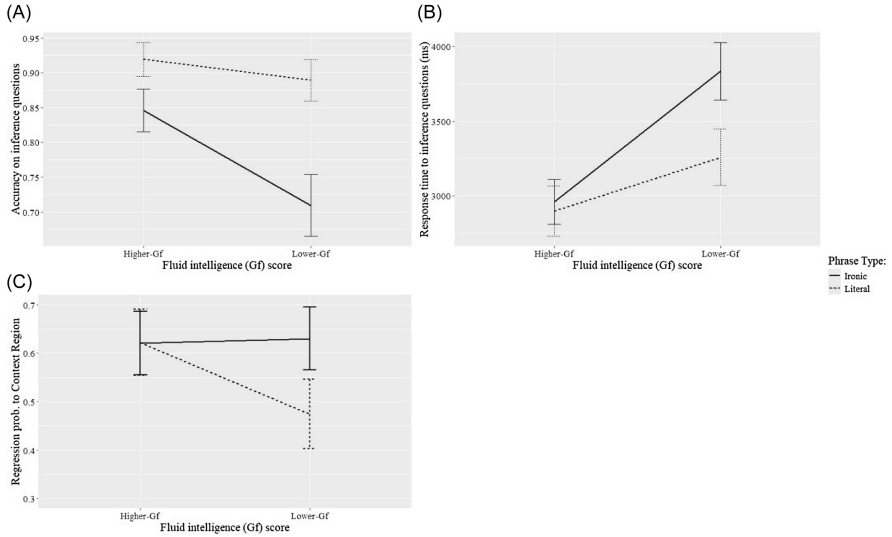


Figure 1. Graphs illustrating interactions between levels of fluid intelligence (Higher-Gf and Lower-Gf) and levels of Phrase Type (Ironic and Literal), for accuracy on inference questions (Panel A), response time to inference questions (Panel B), and regression probability to the context region (Panel C). Error bars represent Standard Error and the means illustrated are the Marginal Means as calculated directly from the models.

interaction between fluid intelligence and Phrase Type in regression probability to the context region ($\beta = -0.01$, $SE = 0.005$, $z = -2.29$, $p = 0.026$): as fluid intelligence increased, regression probability decreased in the irony condition. The same interaction was also approaching significance in the first-pass gaze duration of the target phrase region ($\beta = 0.001$, $SE = 0.001$, $t = 1.91$, $p = 0.056$), whereby as fluid intelligence increased, reading times in the irony condition increased.

To facilitate the interpretation of these data and to conduct further pairwise comparisons, the same models were rerun using a recoded variable for fluid intelligence. Specifically, this predictor was recoded as a factor, namely Raven's bin, with 2 levels: Higher-Gf (where Raven's score $\geq$ mean (Raven's score); $N = 28$) and Lower-Gf (where Raven's score $<$ mean (Raven's score); $N = 29$). Higher-Gf was set as the reference level. Marginal means and model outputs are provided in the appendix (Tables A2 and A3), while main trends are illustrated in Figure 1. Pairwise contrasts and marginal means were estimated using *emmeans()* package (Lenth, 2018), with a Bonferroni p-value adjustment.[2] For brevity, below we only report pairwise contrasts that reached significance and were of interest to the discussion.

For accuracy on inference questions, pairwise contrasts confirmed that Lower-Gf readers were significantly less accurate in the irony as opposed to the literal condition ($p < 0.0001$), while Higher-Gf readers did not reliably differ between the two conditions ($p = 0.12$). Furthermore, Lower-Gf readers were significantly less accurate than Higher-Gf readers in the irony condition ($p = 0.03$), but there was no difference in accuracy between Higher- and Lower-Gf readers in the literal condition ($p = 1.00$; Figure 1, Panel A).

https://doi.org/10.1017/S0142716424000444 Published online by Cambridge University Press

In response time to inference questions, Lower-Gf readers were significantly slower than Higher-Gf readers in the irony condition ($p < 0.001$), but there was no difference between Higher- and Lower-Gf readers in the literal condition ($p = 0.52$; Figure 1, Panel B). For Higher-Gf readers, response time was not significantly modulated by Phrase Type ($p = 1.00$), while Lower-Gf readers were significantly faster in the literal condition ($p < 0.001$).

For regression probability to the context region, pairwise contrasts demonstrated that Lower-Gf readers were significantly more likely to generate a regression in the irony as opposed to the literal condition ($p = 0.03$), while regression probability for Higher-Gf readers was not modulated by Phrase Type ($p = 1.00$; Figure 1, Panel C).

Working memory effects

We did not find any significant effects of WM score in any of the implicit or explicit measures examined.

Discussion

In this study, we investigated whether WM and fluid intelligence, the two main components of executive attention, distinctively modulate the processing of irony, a complex pragmatic phenomenon that often requires the consideration of various contextual factors to infer the intended meaning. In an eye-tracking reading experiment, adult participants read stories containing ironic or literal phrases and answered questions targeting their interpretation. Participants' WM capacity and fluid intelligence were estimated via two separate psychometric tasks, the scores of which were used as predictors in analyses. The results revealed a null effect of WM, while fluid intelligence affected both implicit and explicit measures, indicating involvement in the processing and comprehension of irony. Readers with higher fluid intelligence were significantly faster and more accurate in their responses to inference questions targeting ironic statements, while readers with lower-fluid intelligence exhibited higher regression probability to the context region in the irony as opposed to the literal condition. Before discussing these findings in more detail, we first review general trends pertaining to the processing of irony.

The processing of irony

In line with previous literature, our findings demonstrate that verbal irony incurs a processing cost (e.g., Au-Yeung et al., 2015; Filik et al., 2014, 2018; Kaakinen et al., 2014; Olkoniemi et al., 2016; Olkoniemi, Johander, et al., 2019). Compared to literal phrases, ironic phrases elicited significantly longer reading times in first-pass gaze duration, go-past, and total reading time, and were significantly more likely to yield a regression, indicating a greater need for reanalysis (Conklin et al., 2018). This processing cost may reflect initial or stronger activation of the literal meaning, which then needs to be discarded. Alternatively, it could be attributed to increased mentalizing effort in inferring the speaker's ironic intent.

The findings from the explicit comprehension task paint a similar picture: response times were significantly slower and accuracy significantly lower when

https://doi.org/10.1017/S0142716424000444 Published online by Cambridge University Press

inference questions tapped into the understanding of ironic as opposed to literal utterances. This was despite the high accuracy of the general content questions across both conditions. Accuracy on the inference questions depended on successfully discerning the mental states (e.g., feelings/beliefs) of the fictional story characters who uttered the target phrases, which were not explicitly stated in the text in either the literal or the irony condition. The significant cost observed in the irony condition, therefore, suggests an irony-specific challenge, since deriving similar inferences in the literal condition was evidently easier.

Previous literature suggested that irony is resolved in later stages of processing, as processing costs (if observed) do not occur in early eye-tracking measures, but rather in late ones (e.g., Au-Yeung et al., 2015; Filik et al., 2014, 2018; Filik & Moxey, 2010). In line with this, we did not find an effect of Phrase Type in the first-pass reading time of target phrases. However, we did find an effect in first-pass gaze duration (i.e., duration of all fixations and refixations during first pass), which we still consider an early measure. We therefore take our findings to indicate that irony may be detected and processed earlier than previously assumed.

The (null) effect of working memory

The main purpose of the study was to investigate to what extent executive attention abilities influence the interpretation of irony. We hypothesized that greater WM capacity would lead to earlier irony detection and processing, while lower WM capacity would lead to increased reanalysis, two effects reported in some (but not all) previous studies (Kaakinen et al., 2014; Olkoniemi et al., 2016; Olkoniemi, Johander, et al., 2019). Contrary to our predictions, we did not find a significant effect of WM in any of the measures investigated.

One could argue that the conflicting results in the literature are due to differences in the choice of WM task. Some previous studies, for instance, have used digit (Parola & Bosco, 2022) or reading span tasks (Kaakinen et al., 2014; Olkoniemi et al., 2016), both of which provide an index of verbal WM capacity. In these tasks, participants are required to maintain a sequence of numbers or a word active in memory and then orally repeat it at the end of each trial. Despite the theoretical similarities between digit and reading spans, however, only the latter has so far produced significant effects in the processing of irony (e.g., in Kaakinen et al., 2014; Olkoniemi et al., 2016), yet not consistently across the studies in which it was employed (see Olkoniemi, Strömberg, et al., 2019). The task employed in the present study tapped into spatial as opposed to verbal WM, as participants were required to recall which cells within a matrix turned red and in which order, after performing an (unrelated but cognitively taxing) judgment task and provide their answers by clicking the corresponding boxes in sequence. Given the inconsistencies produced by verbal WM tasks in previous studies, it seems unlikely that our null finding can be explained by our task tapping into spatial WM. In fact, Olkoniemi, Johander et al. (2019) found that scores from a Symmetry Span task, such as the one employed in the current study, were more reliable (and the only significant) predictors in the processing of sarcasm, whereas scores obtained from a (verbal) reading span task were not.

https://doi.org/10.1017/S0142716424000444 Published online by Cambridge University Press

A potential reason accounting for the lack of an SSPAN effect, as opposed to Olkoniemi, Johander et al. (2019), could be the different reading tasks employed. Olkoniemi, Johander et al. (2019) used a mask versus no-mask reading paradigm, in which the text was either replaced by x's once read (mask condition) or remained visible (no-mask condition). Readers with lower (visuospatial) WM were more likely to reread sarcastic utterances and produce shorter first-pass rereading times in the no-mask condition, where the text was still available. Readers with higher (visuospatial) WM, on the other hand, initiated regressions to the location of sarcastic utterances even in the mask condition, when the text was no longer visible. This suggests that readers with higher visuospatial WM were able to use the location of the text as a successful cue for content recall. Given that in our reading task, the text was always visible, SSPAN effects may not have emerged because readers were always able to reread content without having to store text content or its location in WM.

Additionally, it seems unlikely that our null WM effect can be explained by low task demands, as suggested by Olkoniemi, Strömberg, et al. (2019). Participants were required to read 36 stories (including fillers) with an average length of 730 characters, extract and retain key information active in memory, pertaining to the gist of the text, character names, and attitudes as the stories disappeared from the screen ahead of the comprehension task. They were further required to infer characters' attitudes and beliefs based on the interpretation of the target phrases, which were no longer accessible to them, by relying on their WM. We are therefore confident that our reading task was demanding enough to detect potential WM effects.

We cannot of course exclude the possibility that some of the variance attributable to WM may have been captured by our fluid intelligence measure. As we discussed in the Introduction, WM and fluid intelligence are believed to be correlated but separable constructs (Kane et al., 2005; Oberauer et al., 2005), both relying on the same ability to control attention. In the SSPAN task, for example, a participant would need to quickly disengage from the symmetry judgment task (processing component), in order to perform well on the recall part (memory component). In line with this, we found a mild, positive correlation between our WM and fluid intelligence scores, without however evidence of collinearity in the models. Following the suggestions of a reviewer to explore this issue further, we conducted post-hoc mediation analyses using the *lavaan* package (Rosseel, 2012) in R. Specifically, we explored whether fluid intelligence acted as a significant mediator for WM in models where fluid intelligence was significant. The analyses indeed indicated a significant and complete mediation effect on accuracy and response time to inference questions, and regression probability to context (with indirect effects at $ps \leq 0.002$). Conversely, and as expected, direct effects between WM and the dependent variables were not significant in any of the models ($ps \geq 0.07$), in line with the main analyses. These results therefore further support the argument that WM may not exert additional (or direct) influences beyond the common attentional control captured by fluid intelligence. At the same time, these strong mediation effects suggest an indirect involvement. Interestingly, this finding lines up with Martin et al.'s (2020), who also observed an indirect effect of WM through fluid intelligence, in predicting successful reading comprehension (measured as accuracy on passage comprehension questions).

https://doi.org/10.1017/S0142716424000444 Published online by Cambridge University Press

Future studies could further investigate the (direct or indirect) role of WM in irony processing in populations with limited or impaired WM capacity. Such investigations could clarify whether there is a detectable minimal requirement for WM capacity that is essential for processing verbal irony and whether this requirement differs compared to equivalent literal phrases.

The effect of fluid intelligence

To our knowledge, the role of fluid intelligence in the on-line processing of irony has not been previously investigated. We hypothesized that greater fluid intelligence would influence irony processing because of better problem-solving skills and more effective disengagement of the literal meaning, at least when utterances are not by default interpreted ironically. We further hypothesized that fluid intelligence would influence offline responses and late reading measures since problem-solving skills would be mobilized after irony (i.e., the "problem") had been detected. Our results support these predictions. We found that fluid intelligence modulated both accuracy and response time to inference questions, as well as regression probability to the context region.

Readers with lower fluid intelligence were significantly more likely to revisit the context region after encountering an ironic than a literal phrase. In contrast, readers with higher fluid intelligence were unaffected by Phrase Type, although overall regression probability was relatively high. Arguably regressions to the context in readers with lower fluid intelligence reflect compensatory strategies, as part of dealing with the higher demands of irony processing, some of which may emerge from a literal-first interpretation. Conversely, those readers with higher fluid intelligence may reflect strategic processing, whereby regressions in both conditions were used to confirm derived interpretations, to achieve greater accuracy in the comprehension task. It is worth pointing out, however, that even for readers with higher fluid intelligence, *fully* integrating irony was costly as late eye-tracking measures in the target phrase region (i.e., go-past, total reading time, and regression probability) consistently showed a processing cost for irony, without a facilitative (or otherwise) effect of fluid intelligence.

In the offline comprehension task, greater fluid intelligence led to significantly more accurate and significantly faster response times to inference questions in the irony condition, relative to lower fluid intelligence. No differences between higher- and lower-fluid intelligence readers were observed in the literal condition, suggesting that fluid intelligence is particularly important to the comprehension of irony. The comprehension of literal language should not engage problem-solving or disengagement mechanisms to the same degree, since, unlike irony, it does not typically invoke disambiguation processes and hypotheses testing. Except for instances of polysemy or homonymy, for example, literal language does not involve consideration of alternative interpretations. Therefore, arriving at the correct answer to comprehension questions is comparatively easier and faster.

Finally, we found a marginal effect in the first-pass gaze duration of ironic phrases, an early reading measure. Readers with higher fluid intelligence tended to exhibit slower first-pass gaze durations in the irony as opposed to the literal condition, implying that greater fluid intelligence may be involved in the early detection and processing of ironic meanings. However, this finding should be considered with caution given the p-value of the effect ($p = 0.056$).

https://doi.org/10.1017/S0142716424000444 Published online by Cambridge University Press

From the above discussion, it appears that lower fluid intelligence can lead to problems in comprehension of irony, which results in extensive reanalysis of available information. Moreover, when irony was concerned, the accuracy of readers with lower fluid intelligence suffered about 30% of the time, even though they allowed for extra time when responding to inference questions. On the other hand, greater fluid intelligence can improve comprehension accuracy and speed, and lead to more efficient strategic processing.

Returning to our hypothesis, and assuming that fluid intelligence relates to problem-solving skills and the ability to disengage irrelevant information, there are three, not mutually exclusive possibilities that may account for the present data. First, poorer problem-solving skills might make it more difficult for readers to derive or bring alternative solutions (in this case ironic meanings) to the necessary activation threshold. Second, poorer disengagement skills might result in greater activation of irrelevant information, causing interference during processing. Here, interference could stem from over-activating the literal meaning of an ironic utterance, or from an inability to effectively discard it. In general, adequate attention control in reading reduces adverse effects related to mind-wandering, and those resulting from distracting, outdated, and irrelevant information (Christopher et al., 2012; Hasher et al., 1999). Specifically, with insufficient attention control, readers may experience difficulty in blocking irrelevant information (e.g., Borella et al., 2010; De Beni & Palladino, 2000; McVay & Kane, 2012), and nontarget meanings (e.g., Gernsbacher, 1990, 1993; Gernsbacher et al., 1990; Gernsbacher & Faust, 1991), thus directly impacting reading proficiency and skill (Butterfuss & Kendeou, 2018). For instance, even sometime after reading ambiguous words in disambiguating sentences (e.g., *he dug with the spade*, where *spade* can have the meaning of "shovel" (target) or "one of the four suits of playing cards" (nontarget)), less skillful readers experienced sustained interference from nontarget meanings. This was indicated by their inability to judge probe words associated with nontarget meanings (e.g., *ace* for "spade") as semantically irrelevant (Gernsbacher et al., 1990). Our data extend these findings by suggesting that disengagement mechanisms inherent in fluid intelligence are also important to the processing of verbal irony. These mechanisms potentially reduce interference from nontarget literal interpretations of ironic utterances, at least when a literal interpretation is the first one to be accessed.

Third, and in line with our interpretation of the regression analysis, it is possible that readers with lower fluid intelligence resorted to compensatory strategies because of fewer attentional resources. This behavior resembles what has been previously reported for readers with lower WM, who reportedly initiated more lookbacks (i.e., regressions) to ironic than literal phrases (Olkoniemi et al., 2016; Olkoniemi, Johander, et al., 2019). Our findings suggest that fluid intelligence is a more important and reliable predictor in the processing and comprehension of irony, and potentially a mediator for WM. Once this factor is accounted for, variability in WM may not further modulate processing in a direct way. Therefore, our hypothesis that WM and fluid intelligence would both influence irony processing in an independent manner, was only partly confirmed given the null or indirect at best effect of WM.

To better understand the role of executive attention in the processing of irony as well as other pragmatic phenomena, future studies should aim to utilize more comprehensive measures of fluid intelligence and WM. In the present study, we relied

https://doi.org/10.1017/S0142716424000444 Published online by Cambridge University Press

on one task per construct which was necessary to avoid overly long testing times or multiple sessions for participants. Ideally, fluid intelligence and WM capacity should be assessed with multiple tasks to boost construct validity, potentially by extracting latent factors (Draheim et al., 2019). To compensate for this limitation at least partially, we have made sure to select cognitive tasks with attested internal and external validity (Draheim et al., 2021, 2023). It is worth mentioning that modified attention control tasks with good psychometric properties and significantly shorter administration time have recently emerged, which might offer a promising avenue for future research (see Burgoyne et al., 2023).

Conclusion

Our study advances the larger enterprise in experimental pragmatics aiming to map the cognitive and affective foundations of nonliteral language processing, by investigating the cognitive mechanisms involved in the comprehension of irony. Specifically, our findings indicate a significant role of fluid intelligence in the processing of irony in both implicit and explicit comprehension measures, while the role of WM was found to be indirect at best. We argue that fluid intelligence aids the processing of irony by offering better problem-solving skills or better disengagement of literal meanings, allowing for more efficient and accurate processing and integration of ironic interpretations. Based on our findings, future studies on the processing of nonliteral language should consider assessing fluid intelligence as an important individual-difference variable, since it varies with developmental stage and overall cognitive abilities. This might cast new light on why young children and certain clinical populations exhibit specific challenges in understanding irony and other types of nonliteral language.

Replication package. Supplementary Materials including stimuli, data, analysis code, and model outputs are available here: https://osf.io/4f7xm/.

Acknowledgments. We would like to thank Curtis Sharma and David Thornton for proof-reading an earlier version of this article; Cecilie Rummelhoff and Camilla Lee for translating the stimuli stories to Norwegian; and Cecilie Rummelhoff, Fatemeh Montazerikafrani, Valerie Borey, Ellen Margrethe Ulving, and Ola Trondsson Eidet for their help with recruitment and data collection.

Funding. This work was supported by a FINNUT grant from the Research Council of Norway (project no. 315368) awarded to FK.

Competing interests. The author(s) declare none.

Notes

1 When building the experiment, we mistakenly marked an incorrect key as the correct response, resulting in responses always being logged as incorrect. We could not recover the data since key presses and their accuracy were not separately logged. Nonetheless, mixed-effect models are expected to handle the unbalanced dataset effectively.

2 To enable multiple contrasts between Phrase Type and Raven's bin in the accuracy model, we reintroduced the Phrase Type by Raven's bin interaction. Recall that this interaction had been originally dropped in the main analysis due to convergence issues with the full model. Importantly, the follow-up model with Raven's bin produced the same main effects as the final accuracy model of the main analysis.

https://doi.org/10.1017/S0142716424000444 Published online by Cambridge University Press

References

Akimoto, Y., Sugiura, M., Yomogida, Y., Miyauchi, C. M., Miyazawa, S., & Kawashima, R. (2014). Irony comprehension: Social conceptual knowledge and emotional response. *Human Brain Mapping*, **35**(4), 1167–1178.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In *Psychology of learning and motivation* (Vol. **2**, pp. 89–195). Elsevier.

Au-Yeung, S. K., Kaakinen, J. K., Liversedge, S. P., & Benson, V. (2015). Processing of written irony in Autism Spectrum Disorder: An eye-movement study. *Autism Research*, **8**(6), 749–760.

Baddeley, A. (1998). Recent developments in working memory. *Current Opinion in Neurobiology*, **8**(2), 234–238.

Baddeley, A., & Hitch, G. (1974). Working memory. In *Psychology of learning and motivation* (Vol. **8**, pp. 47–89). Elsevier.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. *R Package Version*, **1**(7), 1–23.

Borella, E., Carretti, B., & Pelegrina, S. (2010). The specific role of inhibition in reading comprehension in good and poor comprehenders. *Journal of Learning Disabilities*, **43**(6), 541–552.

Bosco, F. M., Parola, A., Valentini, M. C., & Morese, R. (2017). Neural correlates underlying the comprehension of deceitful and ironic communicative intentions. *Cortex*, **94**, 73–86.

Bruntsch, R., & Ruch, W. (2017). Studying irony detection beyond ironic criticism: Let's include ironic praise. *Frontiers in Psychology*, **8**, 606.

Burgoyne, A. P., & Engle, R. W. (2020). Attention control: A cornerstone of higher-order cognition. *Current Directions in Psychological Science*, **29**(6), 624–630.

Burgoyne, A. P., Jessie, A., Martin, D., Mashburn, C. A., Tsukahara, J. S., Draheim, C., & Engle, R. W. (2022). Measuring individual differences in working memory capacity and attention control and their contribution to language comprehension. In Schwieter, J., Wen, Z., (Eds.) *The Cambridge handbook of working memory and language* (pp. 247–273).

Burgoyne, A. P., Tsukahara, J. S., Mashburn, C. A., Pak, R., & Engle, R. W. (2023). Nature and measurement of attention control. *Journal of Experimental Psychology: General*, **152**(8), 2369–2402. https://doi.org/10.1037/xge0001408

Butterfuss, R., & Kendeou, P. (2018). The role of executive functions in reading comprehension. *Educational Psychology Review*, **30**(3), 801–826.

Caillies, S., Bertot, V., Motte, J., Raynaud, C., & Abely, M. (2014). Social cognition in ADHD: Irony understanding and recursive theory of mind. *Research in Developmental Disabilities*, **35**(11), 3191–3198. https://doi.org/10.1016/j.ridd.2014.08.002

Carretti, B., Borella, E., Cornoldi, C., & De Beni, R. (2009). Role of working memory in explaining the performance of individuals with specific reading comprehension difficulties: A meta-analysis. *Learning and Individual Differences*, **19**(2), 246–251.

Cattell, R. B. (1943). The description of personality. I. Foundations of trait measurement. *Psychological Review*, **50**(6), 559.

Christopher, M. E., Miyake, A., Keenan, J. M., Pennington, B., DeFries, J. C., Wadsworth, S. J., Willcutt, E., & Olson, R. K. (2012). Predicting word reading and comprehension with executive function and speed measures across development: a latent variable analysis. *Journal of Experimental Psychology: General*, **141**(3), 470.

Clark, H. H., & Gerrig, R. J. (1984). On the pretense theory of irony. *Journal of Experimental Psychology: General*, **113**(1), 121–126.

Conklin, K., Pellicer-Sánchez, A., & Carrol, G. (2018). *Eye-Tracking: A Guide for Applied Linguistics Research*. Cambridge University Press.

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, **12**, 769–786.

Cornejol, C., Simonetti, F., Aldunate, N., Ibáñez, A., López, V., & Melloni, L. (2007). Electrophysiological evidence of different interpretative strategies in irony comprehension. *Journal of Psycholinguistic Research*, **36**(6), 411–430.

https://doi.org/10.1017/S0142716424000444 Published online by Cambridge University Press

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, **19**(4), 450–466.

De Beni, R., & Palladino, P. (2000). Intrusion errors in working memory tasks: Are they related to reading comprehension ability? *Learning and Individual Differences*, **12**(2), 131–143.

Dews, S., & Winner, E. (1999). Obligatory processing of literal and nonliteral meanings in verbal irony. *Journal of Pragmatics*, **31**(12), 1579–1599.

Draheim, C., Mashburn, C. A., Martin, J. D., & Engle, R. W. (2019). Reaction time in differential and developmental research: A review and commentary on the problems and alternatives. *Psychological Bulletin*, **145**(5), 508.

Draheim, C., Tshukara, J. S., & Engle, R. W. (2023). Replication and extension of the toolbox approach to measuring attention control. *Behavior Research Methods*, **56**, 1–23.

Draheim, C., Tsukahara, J. S., Martin, J. D., Mashburn, C. A., & Engle, R. W. (2021). A toolbox approach to improving the measurement of attention control. *Journal of Experimental Psychology: General*, **150**(2), 242.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, **11**(1), 19–23.

Engle, R. W. (2018). Working memory and executive attention: A revisit. *Perspectives on Psychological Science*, **13**(2), 190–193.

Falkum, I. L., & Köder, F. (2020). The acquisition of figurative meanings. *Journal of Pragmatics*, **164**, 18–24. https://doi.org/10.1016/j.pragma.2020.04.007

Filik, R., Howman, H., Ralph-Nearman, C., & Giora, R. (2018). The role of defaultness and personality factors in sarcasm interpretation: Evidence from eye-tracking during reading. *Metaphor and Symbol*, **33**(3), 148–162.

Filik, R., Leuthold, H., Wallington, K., & Page, J. (2014). Testing theories of irony processing using eye-tracking and ERPs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **40**(3), 811.

Filik, R., & Moxey, L. M. (2010). The on-line processing of written irony. *Cognition*, **116**(3), 421–436.

Filippova, E., & Astington, J. W. (2008). Further development in social reasoning revealed in discourse irony understanding. *Child Development*, **79**(1), 126–138.

Foster, J. L., Shipstead, Z., Harrison, T. L., Hicks, K. L., Redick, T. S., & Engle, R. W. (2015). Shortened complex span tasks can reliably measure working memory capacity. *Memory & Cognition*, **43**(2), 226–236.

Friedman, N. P., & Miyake, A. (2017). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex*, **86**, 186–204.

Gernsbacher, M. A. (1990). *Language comprehension as structure building*. Erlbaum.

Gernsbacher, M. A. (1993). *Less skilled readers have less efficient suppression mechanisms*. SAGE Publications Sage.

Gernsbacher, M. A., & Faust, M. E. (1991). The mechanism of suppression: a component of general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **17**(2), 245.

Gernsbacher, M. A., Varner, K. R., & Faust, M. E. (1990). Investigating differences in general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **16**(3), 430.

Gibbs, R. W. (1994). *The poetics of mind: Figurative thought, language, and understanding*. Cambridge University Press.

Gibbs, R. W. (2000). Irony in talk among friends. *Metaphor and Symbol*, **15**(1–2), 5–27.

Giora, R. (2003). *On our mind: Salience, context, and figurative language*. Oxford University Press.

Giora, R., Fein, O., & Schwartz, T. (1998). Irony: grade salience and indirect negation. *Metaphor and Symbol*, **13**(2), 83–101.

Giora, R., Givoni, S., & Fein, O. (2015). Defaultness reigns: The case of sarcasm. *Metaphor and Symbol*, **30**(4), 290–313.

Godbee, K., & Porter, M. (2013). Comprehension of sarcasm, metaphor and simile in Williams syndrome. *International Journal of Language & Communication Disorders*, **48**(6), 651–665.

Grice, H. P. (1975). Logic and conversation. In *Speech acts* (pp. 41–58). Brill.

Hasher, L., Zacks, R. T., & May, C. P. (1999). *Inhibitory control, circadian arousal, and age*.

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology*, **57**(5).

https://doi.org/10.1017/S0142716424000444 Published online by Cambridge University Press

Jacob, H., Kreifelts, B., Nizielski, S., Schütz, A., & Wildgruber, D. (2016). Effects of emotional intelligence on the impression of irony created by the mismatch between verbal and nonverbal cues. *PloS One*, **11**(10), e0163211.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, **59**(4), 434–446.

Kaakinen, J. K., Olkoniemi, H., Kinnari, T., & Hyönä, J. (2014). Processing of written irony: An eye movement study. *Discourse Processes*, **51**(4), 287–311.

Kałowski, P., Zajączkowska, M., Branowska, K., Olechowska, A., Siemieniuk, A., Dryll, E., & Banasik-Jemielniak, N. (2023). Individual differences in verbal irony use: A systematic review of quantitative psycholinguistic studies. *Metaphor and Symbol*, **38**(1), 81–111.

Kane, M. J., Hambrick, D. Z., & Conway, A. R. A. (2005). *Working memory capacity and fluid intelligence are strongly related constructs: comment on Ackerman, Beier, and Boyle (2005).*

Katz, A. N., Blasko, D. G., & Kazmerski, V. A. (2004). Saying what you don't mean: Social influences on sarcastic language processing. *Current Directions in Psychological Science*, **13**(5), 186–189.

Katz, A. N., & Ferretti, T. R. (2001). Moment-by-moment reading of proverbs in literal and nonliteral contexts. *Metaphor and Symbol*, **16**(3–4), 193–221.

Kreuz, R. J., & Link, K. E. (2002). Asymmetries in the use of verbal irony. *Journal of Language and Social Psychology*, **21**(2), 127–143.

Kumle, L., Võ, M. L.-H., & Draschkow, D. (2021). Estimating power in (generalized) linear mixed models: An open introduction and tutorial in R. *Behavior Research Methods*, **53**(6), 2528–2543.

Kumle, L., Võ, M. L. H., & Draschkow, D. (2018). Mixedpower: a library for estimating simulation-based power for mixed models in R. Computer Software]. https://Github.Com/DejanDraschkow/Mixedpower.

Lenth, R. (2018). Emmeans: Estimated marginal means, aka least-square means. R Package Version 1.2.

Martin, I., & McDonald, S. (2005). Evaluating the causes of impaired irony comprehension following traumatic brain injury. *Aphasiology*, **19**(8), 712–730. https://doi.org/10.1080/02687030500172203

Martin, J. D., Shipstead, Z., Harrison, T. L., Redick, T. S., Bunting, M., & Engle, R. W. (2020). The role of maintenance and disengagement in predicting reading comprehension and vocabulary learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **46**(1), 140.

Mashburn, C. A., Burgoyne, A. P., & Engle, R. W. (2023). Working memory, intelligence, and life success. In *Memory in science for society: There is nothing as practical as a good theory* (p. 149). Oxford University Press.

McVay, J. C., & Kane, M. J. (2012). Why does working memory capacity predict variation in reading comprehension? On the influence of mind wandering and executive attention. *Journal of Experimental Psychology: General*, **141**(2), 302.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, **41**(1), 49–100.

Nicholson, A., Whalen, J. M., & Pexman, P. M. (2013). Children's processing of emotion in ironic language. *Frontiers in Psychology*, **4**, 691.

Nilsen, E. S., Glenwright, M., & Huyder, V. (2011). Children and adults understand that verbal irony interpretation depends on listener knowledge. *Journal of Cognition and Development*, **12**(3), 374–409.

Oberauer, K., Schulze, R., Wilhelm, O., & Süß, H.-M. (2005). Working memory and intelligence–their correlation and their relation: comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, **131**(1), 61–65.

Obert, A., Gierski, F., Calmus, A., Flucher, A., Portefaix, C., Pierot, L., Kaladjian, A., & Caillies, S. (2016). Neural correlates of contrast and humor: processing common features of verbal irony. *PLoS One*, **11**(11), e0166704.

Olkoniemi, H., Johander, E., & Kaakinen, J. K. (2019). The role of look-backs in the processing of written sarcasm. *Memory & Cognition*, **47**(1), 87–105.

Olkoniemi, H., & Kaakinen, J. K. (2021). Processing of irony in text: A systematic review of eye-tracking studies. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, **75**(2), 99–106.

Olkoniemi, H., Ranta, H., & Kaakinen, J. K. (2016). Individual differences in the processing of written sarcasm and metaphor: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **42**(3), 433.

https://doi.org/10.1017/S0142716424000444 Published online by Cambridge University Press

Olkoniemi, H., Strömberg, V., & Kaakinen, J. K. (2019). The ability to recognise emotions predicts the time-course of sarcasm processing: Evidence from eye movements. *Quarterly Journal of Experimental Psychology*, **72**(5), 1212–1223.

Paap, K. R., Johnson, H. A., & Sawi, O. (2015). Bilingual advantages in executive functioning either do not exist or are restricted to very specific and undetermined circumstances. *Cortex*, **69**, 265–278.

Parola, A., & Bosco, F. M. (2022). An eye-tracking investigation of the cognitive processes involved in the comprehension of simple and complex communicative acts. *Quarterly Journal of Experimental Psychology*, **75**(10), 1976–1995.

Pearson. (2015). *Raven's advanced progressive matrices (APM-III): manual part I*.

Pexman, P. M. (2008). It's fascinating research: The cognition of verbal irony. *Current Directions in Psychological Science*, **17**(4), 286–290.

Posner, M. I., & DiGirolamo, G. J. (1998). Conflict, target detection and cognitive control. In R. Parasuraman (Ed.), *The attentive brain* (pp. 401–423).

R Core Team. (2023). R: A language and environment for statistical computing. http://www.r-project.org/

Recchia, H. E., Howe, N., Ross, H. S., & Alexander, S. (2010). Children's understanding and production of verbal irony in family conversations. *British Journal of Developmental Psychology*, **28**(2), 255–274. https://doi.org/10.1348/026151008X401903

Regel, S., Coulson, S., & Gunter, T. C. (2010). The communicative style of a speaker can affect language comprehension? ERP evidence from the comprehension of irony. *Brain Research*, **1311**, 121–135.

Regel, S., Gunter, T. C., & Friederici, A. D. (2011). Isn't it ironic? An electrophysiological exploration of figurative language processing. *Journal of Cognitive Neuroscience*, **23**(2), 277–293.

Rey-Mermet, A., Gade, M., & Oberauer, K. (2018). Should we stop thinking about inhibition? Searching for individual and age differences in inhibition ability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **44**(4), 501.

Ronderos, C. R., Tomlinson, J. M., & Noveck, I. (2023). When irony is faster than its literal control: The role of mindreading during irony comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **50**(3), 509–522.

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, **48**(2 SE-Articles), 1–36. https://doi.org/10.18637/jss.v048.i02

Rouder, J. N., & Haaf, J. M. (2019). A psychometrics of individual differences in experimental tasks. *Psychonomic Bulletin & Review*, **26**(2), 452–467.

Shibata, M., Toyomura, A., Itoh, H., & Abe, J. (2010). Neural substrates of irony comprehension: A functional MRI study. *Brain Research*, **1308**, 114–123.

Shin, J. (2020). A meta-analysis of the relationship between working memory and second language reading comprehension: Does task type matter? *Applied Psycholinguistics*, **41**(4), 873–900.

Shipstead, Z., Harrison, T. L., & Engle, R. W. (2016). Working memory capacity and fluid intelligence: Maintenance and disengagement. *Perspectives on Psychological Science*, **11**(6), 771–799.

Sperber, D., & Wilson, D. (1981). Irony and the use-mention distinction. *Philosophy*, **3**, 143–184.

Spotorno, N., Cheylus, A., Van Der Henst, J.-B., & Noveck, I. A. (2013). What's behind a P600? Integration operations during irony processing. *PLoS One*, **8**(6), e66839.

Spotorno, N., Koun, E., Prado, J., Van Der Henst, J.-B., & Noveck, I. A. (2012). Neural evidence that utterance-processing entails mentalizing: The case of irony. *NeuroImage*, **63**(1), 25–39.

Spotorno, N., & Noveck, I. A. (2014). When is irony effortful? *Journal of Experimental Psychology: General*, **143**(4), 1649.

Tsukahara, J. S. (2022). *englelab: An R package for processing complex-span and attention control tasks downloaded from the EngleLab (1.1.0)*. Zenodo.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, **28**(2), 127–154.

Utsumi, A. (2000). Verbal irony as implicit display of ironic environment: Distinguishing ironic utterances from nonirony. *Journal of Pragmatics*, **32**(12), 1777–1806.

Wang, A. T., Lee, S. S., Sigman, M., & Dapretto, M. (2006). Neural basis of irony comprehension in children with autism: The role of prosody and context. *Brain*, **129**(4), 932–943. https://doi.org/10.1093/brain/awl032

Wilson, D., & Sperber, D. (2012). Explaining irony. In D. Wilson & D. Sperber (Eds.), *Meaning and relevance* (pp. 123–145). Cambridge University Press.

https://doi.org/10.1017/S0142716424000444 Published online by Cambridge University Press

https://doi.org/10.1017/S0142716424000444 Published online by Cambridge University Press

Appendix

Table A1. Model outputs of main analyses

	TARGET PHRASE REGION											
	First pass reading time				First pass gaze duration				Go-past reading time			
Predictors	*β*	*SE*	*t*	*p*	*β*	*SE*	*t*	*p*	*β*	*SE*	*t*	*p*
(Intercept)	6.546	0.064	102.211	**<0.001**	6.721	0.072	93.582	**<0.001**	6.781	0.072	94.335	**<0.001**
PhraseType[Irony]	0.006	0.014	0.43	0.667	0.02	0.01	1.985	**0.047**	0.032	0.012	2.561	**0.011**
z Raven's score	−0.004	0.002	−1.484	0.138	−0.004	0.003	−1.502	0.133	−0.003	0.003	−1.305	0.192
z SSPAN score	0.002	0.005	0.484	0.629	0.004	0.005	0.867	0.386	0.005	0.005	0.957	0.339
z Trial Index	−0.002	0.002	−0.866	0.387	−0.006	0.001	−4.308	**<0.001**	−0.006	0.002	−3.75	**<0.001**
PhraseType[Irony] × z Raven's score	0.001	0.001	1.505	0.132	0.001	0.001	1.91	0.056	0.001	0.001	0.967	0.333
PhraseType[Irony] × z SSPAN score	−0.002	0.002	−0.925	0.355	-0.001	0.001	−0.425	0.671	0.001	0.001	0.471	0.638
Random Effects												
σ^2	0.26				0.12				0.15			
τ_{00}	$0.06_{\text{ppt_id}}$				$0.07_{\text{ppt_id}}$				$0.08_{\text{ppt_id}}$			
	$0.07_{\text{item_id}}$				$0.09_{\text{item_id}}$				$0.09_{\text{item_id}}$			
τ_{11}					$0.00_{\text{item_id. PhraseType[Irony]}}$				$0.00_{\text{ppt_id. PhraseType[Irony]}}$			
									$0.00_{\text{item_id. PhraseType[Irony]}}$			
ρ_{01}					$0.16_{\text{item_id}}$				$-0.11_{\text{ppt_id}}$			
									$0.39_{\text{item_id}}$			
ICC	0.33				0.59				0.52			
N	$57_{\text{ppt_id}}$				$57_{\text{ppt_id}}$				$57_{\text{ppt_id}}$			

(*Continued*)

https://doi.org/10.1017/S0142716424000444 Published online by Cambridge University Press

Table A1. (*Continued*)

	TARGET PHRASE REGION											
	First pass reading time				First pass gaze duration				Go-past reading time			
Predictors	*β*	*SE*	*t*	*p*	*β*	*SE*	*t*	*p*	*β*	*SE*	*t*	*p*
	24 $_{item_id}$				24 $_{item_id}$				24 $_{item_id}$			
Observations	1340				1361				1361			
Marginal R^2/Conditional R^2	0.009/0.335				0.020/0.597				0.018/0.532			

	TARGET PHRASE REGION							
	Total reading time				Regression probability			
Predictors	*β*	*SE*	*t*	*p*	*β*	*SE*	*z*	*p*
(Intercept)	6.866	0.07	97.962	**<0.001**	−1.458	0.144	−10.104	**<0.001**
PhraseType[Irony]	0.039	0.014	2.755	**0.006**	0.217	0.08	2.727	**0.006**
z Raven's score	−0.003	0.003	−0.896	0.37	0.004	0.009	0.476	0.634
z SSPAN score	0.002	0.005	0.451	0.652	0.009	0.017	0.519	0.604
z Trial Index	−0.006	0.002	−3.734	**<0.001**	-0.002	0.01	−0.217	0.828
PhraseType[Irony] × z Raven's score	0.001	0.001	1.509	0.132	−0.005	0.005	−0.944	0.345
PhraseType[Irony] × z SSPAN score	0.001	0.001	0.506	0.613	0.014	0.009	1.505	0.132
Random Effects								
σ^2	0.15				3.29			
τ_{00}	0.09 $_{ppt_id}$				0.65 $_{ppt_id}$			
	0.08 $_{item_id}$				0.07 $_{item_id}$			

(*Continued*)

https://doi.org/10.1017/S0142716424000444 Published online by Cambridge University Press

Table A1. (*Continued*)

	TARGET PHRASE REGION							
	Total reading time				Regression probability			
Predictors	*β*	*SE*	*t*	*p*	*β*	*SE*	*z*	*p*
τ_{11}	$0.00_{\text{item_id. PhraseType[Irony]}}$				$0.01_{\text{ppt_id. PhraseType[Irony]}}$			
ρ_{01}	$0.17_{\text{item_id}}$				$-0.11_{\text{ppt_id}}$			
ICC	0.53				0.18			
N	$57_{\text{ppt_id}}$				$57_{\text{ppt_id}}$			
	$24_{\text{item_id}}$				$24_{\text{item_id}}$			
Observations	1351				1328			
Marginal R^2/Conditional R^2	0.015/0.541				0.018/0.196			

	SPILLOVER REGION											
	First pass reading time				First pass gaze duration				Total reading time			
Predictors	*β*	*SE*	*t*	*p*	*β*	*SE*	*t*	*p*	*β*	*SE*	*t*	*p*
(Intercept)	7.906	0.071	111.817	**<0.001**	8.387	0.067	125.466	**<0.001**	8.387	0.067	125.495	**<0.001**
PhraseType[Irony]	−0.091	0.025	−3.675	**<0.001**	0.004	0.007	0.608	0.543	0.004	0.007	0.585	0.559
z Raven's score	0	0.003	0.012	0.99	−0.003	0.003	−1.22	0.223	−0.003	0.003	−1.211	0.226
z SSPAN score	0.004	0.006	0.617	0.537	0.002	0.005	0.328	0.743	0.002	0.005	0.327	0.743
z Trial Index	−0.005	0.004	−1.438	0.151	−0.004	0.001	−4.241	**<0.001**	−0.004	0.001	−4.265	**<0.001**
PhraseType[Irony] × z Raven's score	0	0.002	0.102	0.918	0	0	0.274	0.784	0	0	0.271	0.786
PhraseType[Irony] × z SSPAN score	−0.001	0.003	−0.28	0.78	0	0.001	0.198	0.843	0	0.001	0.2	0.841

(*Continued*)

https://doi.org/10.1017/S0142716424000444 Published online by Cambridge University Press

Table A1. (*Continued*)

	SPILLOVER REGION											
	First pass reading time				First pass gaze duration				Total reading time			
Predictors	*β*	*SE*	*t*	*p*	*β*	*SE*	*t*	*p*	*β*	*SE*	*t*	*p*
Random Effects												
σ^2	0.84				0.05				0.05			
τ_{00}	$0.08_{\text{ppt_id}}$				$0.08_{\text{ppt_id}}$				$0.08_{\text{ppt_id}}$			
	$0.07_{\text{item_id}}$				$0.07_{\text{item_id}}$				$0.07_{\text{item_id}}$			
τ_{11}					$0.00_{\text{ppt_id. PhraseType[Irony]}}$				$0.00_{\text{ppt_id. PhraseType[Irony]}}$			
									$0.00_{\text{item_id. PhraseType[Irony]}}$			
ρ_{01}					$-0.48_{\text{ppt_id}}$				$-0.47_{\text{ppt_id}}$			
									$0.67_{\text{item_id}}$			
ICC	0.15				0.75				0.75			
N	$57_{\text{ppt_id}}$				$57_{\text{ppt_id}}$				$57_{\text{ppt_id}}$			
	$24_{\text{item_id}}$				$24_{\text{item_id}}$				$24_{\text{item_id}}$			
Observations	1362				1362				1362			
Marginal R^2/Conditional R^2	0.011/0.162				0.014/0.756				0.014/0.757			

	CONTEXT REGION				COMPREHENSION TASK [Inference question]							
	Regression probability				Accuracy				Response time			
Predictors	*β*	*SE*	*z*	*p*	*β*	*SE*	*z*	*p*	*β*	*SE*	*t*	*p*
(Intercept)	0.353	0.182	1.944	0.052	1.79	0.18	9.926	**<0.001**	8.078	0.042	193.545	**<0.001**
PhraseType[Irony]	0.161	0.089	1.803	0.071	−0.502	0.107	−4.698	**<0.001**	0.047	0.015	3.194	**0.001**
z Raven's score	0.02	0.012	1.638	0.101	0.028	0.009	3.071	**0.002**	−0.005	0.002	−2.514	**0.012**
z SSPAN score	−0.008	0.023	−0.37	0.711	−0.03	0.017	−1.785	0.074	0.003	0.004	0.924	0.356

(*Continued*)

https://doi.org/10.1017/S0142716424000444 Published online by Cambridge University Press

Table A1. (*Continued*)

	CONTEXT REGION				COMPREHENSION TASK [Inference question]							
	Regression probability				Accuracy				Response time			
Predictors	*β*	*SE*	*z*	*p*	*β*	*SE*	*z*	*p*	*β*	*SE*	*t*	*p*
z Trial Index	−0.03	0.009	−3.183	**0.001**	0.032	0.013	2.578	**0.01**	−0.013	0.001	−8.783	**<0.001**
PhraseType[Irony] × z Raven's score	−0.011	0.005	−2.228	**0.026**					−0.001	0.001	−1	0.318
PhraseType[Irony] × z SSPAN score	0	0.009	−0.05	0.96					0.001	0.002	0.606	0.545
Random Effects												
σ^2	3.29				3.29				0.14			
$\tau_{00\ ppt_id}$	1.48				0.49 $_{ppt_id}$				0.04 $_{ppt_id}$			
$\tau_{00\ item_id}$	0.06				0.24 $_{item_id}$				0.02 $_{item_id}$			
$\tau_{11\ ppt_id.\ PhraseType[Irony]}$	0.05								0.00 $_{ppt_id.\ PhraseType[Irony]}$			
$\tau_{11\ item_id.\ PhraseType[Irony]}$	0.07								0.00 $_{item_id.\ PhraseType[Irony]}$			
$\rho_{01\ ppt_id}$	−0.32								−0.05 $_{ppt_id}$			
$\rho_{01\ item_id}$	0.83								−0.45 $_{item_id}$			
ICC	0.34				0.18				0.33			
N $_{ppt_id}$	57				57 $_{ppt_id}$				57 $_{ppt_id}$			
N $_{item_id}$	24				24 $_{item_id}$				24 $_{item_id}$			
Observations	1326				1000				1357			
Marginal R^2/Conditional R^2	0.035/0.358				0.098/0.262				0.071/0.377			

https://doi.org/10.1017/S0142716424000444 Published online by Cambridge University Press

Table A2. Model outputs of additional models run with Raven's score as a two-level factor (lower and higher fluid intelligence)

	Accuracy [Inference question]				Response time [Inference question]				Regression probability to Context Region			
Predictors	*β*	*SE*	*z*	*p*	*β*	*SE*	*t*	*p*	*β*	*SE*	*z*	*p*
(Intercept)	2.071	0.242	8.563	**<0.001**	7.982	0.05	158.42	**<0.001**	0.497	0.261	1.901	0.057
PhraseType[Irony]	−0.363	0.158	−2.305	**0.021**	0.011	0.019	0.567	0.571	−0.003	0.115	−0.028	0.978
Raven'sBin [Lower-Gf]	−0.578	0.291	−1.986	**0.047**	0.188	0.057	3.278	**0.001**	−0.281	0.365	−0.77	0.441
z SSPAN score	−0.022	0.019	−1.183	0.237	0.004	0.004	1.069	0.285	−0.003	0.023	−0.124	0.901
z Trial Index	0.033	0.013	2.628	**0.009**	−0.013	0.001	−8.845	**<0.001**	−0.03	0.009	−3.201	**0.001**
PhraseType[Irony] × Raven'sBin [Lower-Gf]	−0.231	0.207	−1.115	0.265	0.071	0.025	2.837	**0.005**	0.321	0.147	2.18	**0.029**
PhraseType[Irony] × z SSPAN score	−0.009	0.013	−0.672	0.502	0.002	0.002	1.06	0.289	−0.001	0.009	−0.095	0.924
Random Effects												
σ^2	3.29				0.14				3.29			
τ_{00}	$0.53_{\text{ppt_id}}$				$0.04_{\text{ppt_id}}$				$1.52_{\text{ppt_id}}$			
	$0.25_{\text{item_id}}$				$0.02_{\text{item_id}}$				$0.06_{\text{item_id}}$			
τ_{11}					$0.00_{\text{ppt_id. PhraseType[Irony]}}$				$0.05_{\text{ppt_id. PhraseType[Irony]}}$			
					$0.00_{\text{item_id. PhraseType[Irony]}}$				$0.06_{\text{item_id. PhraseType[Irony]}}$			
ρ_{01}					$-0.29_{\text{ppt_id}}$				$-0.36_{\text{ppt_id}}$			
					$-0.46_{\text{item_id}}$				$0.83_{\text{item_id}}$			
ICC	0.19				0.32				0.34			
N	$57_{\text{ppt_id}}$				$57_{\text{ppt_id}}$				$57_{\text{ppt_id}}$			
	$24_{\text{item_id}}$				$24_{\text{item_id}}$				$24_{\text{item_id}}$			
Observations	1000				1357				1326			
Marginal R^2/Conditional R^2	0.083/0.258				0.089/0.376				0.021/0.354			

https://doi.org/10.1017/S0142716424000444 Published online by Cambridge University Press

Table A3. Marginal Means and Standard Errors of model outputs run with Raven's score as a two-level factor (lower- and higher fluid intelligence)

	Higher fluid intelligence				Lower fluid intelligence			
	Literal		Ironic		Literal		Ironic	
	Mean	*SE*	Mean	*SE*	Mean	*SE*	Mean	*SE*
Comprehension task								
Accuracy	0.92	0.02	0.85	0.03	0.89	0.03	0.71	0.04
Response time	2897	199.9	2960	173.5	3257	222.3	3835	222.7
Context region								
Regression probability	0.62	0.07	0.62	0.07	0.47	0.07	0.63	0.06

Cite this article: Kyriacou, M. & Köder, F. (2024). The cognitive underpinnings of irony comprehension: Fluid intelligence but not working memory modulates processing. *Applied Psycholinguistics* **45**, 1219–1250. https://doi.org/10.1017/S0142716424000444