The Little Grey Cells

by James Jackson

This year marks the centenary of the publication of Agatha Christie's first novel, *The Mysterious Affair at Styles*. In the second chapter she introduces her iconic Belgian detective, Monsieur Hercule Poirot: "an extraordinary-looking little man", with a head "exactly the shape of an egg" and a moustache that "was very stiff and military". Over five decades, Agatha Christie (1890-1976) wrote an impressive sixty-six detective novels - exactly half feature Poirot, and twelve feature her equally renowned spinster-cum-sleuth, Miss Marple, famous inhabitant of St. Mary Mead.

A hundred years on, it is safe to say that her whodunits remain as popular as ever. Barely a year seems to pass without some new adaptation appearing on the big or small screen. Just look at recent Poirots: there's been *Murder on the Orient Express* starring Kenneth Branagh in the lead role, *The A.B.C. Murders* with John Malkovich, and it is less than ten years since the excellent David Suchet drew the *Curtain* on the role after exhaustively covering the stories in ITV's *Agatha Christie's Poirot*. Up-and-coming Christie adaptations include *Death on the Nile*, again starring Branagh. Whilst, the BBC are adapting *Death Comes As The End*; set in Ancient Egypt, the novel is an early example of the historical whodunit, a sub-genre that includes the Cadfael series by Ellis Peters (the *nom de plume* of Edith Pargeter, 1913-1995) and the ongoing Shardlake series by C.J. Sansom.

Agatha Christie began publishing novels during the advent of the whodunit, early on in the period now regularly termed the "Golden Age of Detective Fiction" [1]. Roughly the interval between the two world wars, this is the period when whodunits experienced a surge in popularity. Many of the writers were, notably, women, such as Dorothy L. Sayers (1893-1957), Margery Allingham (1904-1966), Ngaio Marsh (1895-1982) and Josephine Tey (1896-1952). To this day, whodunits remain synonymous with this period - the glamour of the 1920s and 30s - and this nostalgic element undoubtedly plays a part in their continued appeal.

Another reason for the whodunit's popularity is the opportunity for the reader to pit their wits against the detective. With Christie, this inevitably leaves the reader feeling foolish when, in the denouement, Poirot assembles the suspects together and dramatically roots out the murderer. Christie's solutions come with a shock! Perhaps the culprit was the one person you did not suspect. Surely her mysteries are unsolvable. Or are they?

Statistical analysis

In 2015, a study commissioned by UKTV channel, *Drama*, looked at a sample of novels and tried to discover indicators that could help to identify the killer [2]. Specifically, they looked at the language used and the plot structure. This led to a series of interesting findings for example, a female killer is likely if there are "several land vehicles in the story" and a male killer is likely if the victim is strangled. Are there underlying patterns within her stories?

Data and model

Inspired by this previous work, to look at whether information about the characters can help to reveal the culprit, characters were analysed across fifty-six novels. Some books were deemed to stray too far away from the standard whodunit format; for example, *Passenger to Frankfurt* is more of a spy novel; *They Came to Baghdad* is more of an adventure novel.

Three attributes were identified about each character: their gender, whether they are a family relative to any of the victims and their occupation. Websites such as https://knowingchristie.wordpress.com./ proved invaluable in helping to establish characters' occupations. These were pigeonholed into categories that included classic whodunit roles such as "Butler/Housekeeper" and "Clergymen", and also more generic titles such as "Arts", which encompassed a range of professions such as actors, artists and writers. Typically, each character belonged to just one category, which meant there were many zero cells in the data frame - many zero "little grey cells"!

Two types of characters were not considered as suspects. The first of these were the murder victims themselves; incidentally, this meant there were no suspects left in *And Then There Were None*! The second type were the recurring characters, such as Poirot's loyal sidekick, Captain Hastings, and Miss Marple's nephew, Raymond West. Every other character, were treated as suspects and given equal weighting in the analysis.

The response variable, whether or not a character was a killer, was binary. Therefore, an appropriate model to fit was a logistic regression model, which estimates the effect of predictors on the (log) odds of an event. However, a limitation with this model is the lack of independence between the observations within a novel, as knowing that one character is the murderer usually means that the others are not. This dependence has been ignored in the analysis. After all, multiple-murderer cases do occur with one very well-known example – set on a train!

Marginal likelihood ratio tests were carried out to evaluate the relationship between a character's occupation and the odds of them being the murderer. The categorical covariate describing occupation was broken down into a series of indicator variables. The effect of each occupation was then assessed in turn, by comparing the model fit when

the occupation was included in the model, to the model fit when not, that is, to the null model with only an intercept term. The improvement in the model fit was assessed by a likelihood ratio test, which assumes that the difference in the log-likelihoods of two nested models, doubled, is χ^2 distributed. This distribution had one degree of freedom, since there was one extra parameter in the more complex model.

An occupation was deemed to be "significant" if the *p*-value from this test - that is, the probability of observing a result at least as extreme given that the less complex model is the true model (null hypothesis) – fell below 0.05.

Results

There were two occupations for which a "significant" association was uncovered. For members of the police, the association with being the murderer was negative (*p*-value of 0.004) - readers had better look elsewhere.

The policeman's role tends to be ably (or unably) assisting the detective in the investigation - or in the absence of Poirot and Marple - investigating the crime themselves. This negative association could demonstrate that Christie plays fair with her readers, and follows rule number seven of the "decalogue" of detective stories: that the "detective must not himself commit the crime". These rules are a fun set of instructions to encourage detective story writers to play fair with readers. They were devised by Roman Catholic priest Monsignor Ronald Knox (1888-1957), who himself dabbled in detective story writing, and were included in the preface to *Best Detective Stories of 1928-29*.

In contrast, "aristocrats and wealthy" characters had a positive association with being the murderer (*p*-value of 0.016). It appears that the cliché, "the butler did it" - a phrase coined in relation to Mary Roberts Rinehart's 1930 novel *The Door* - needs to be modified to "the butler's boss did it"!

Two other covariates were tested. Whereas the first of these, the character's gender, showed almost no association (*p*-value of 0.94), near independence, the second, whether the character is a family relative to any of the victims, displayed a strong positive association (*p*-value of 0.019) - it seems that readers should not look far from the victim's home for their killer.

A downside of having a relatively large number of occupations is that the robustness of the model is weak. For example, in the books considered, there was one case of a policeman being the murderer. If policemen were considered as a category in their own right, then the estimated regression coefficient would be so low that the model essentially rules that policemen are never the killer. To undertake more robust tests – and to build a model more suited to prediction - the number of categories was collapsed

by grouping similar occupations. A list of these categories, along with a couple of examples of each are shown in Table 1.

A logistic regression model was fitted with the collapsed occupation variables and the "relative" variable as covariates. The conditional odds ratios are presented in Table 1.

For example, the odds ratio of 1.82 for servants means that the odds of a servant being the murderer are 82% greater than for a character with no obvious occupation (assuming that the values of the other covariates remain unchanged).

For a fun exercise, the novel *Lord Edgware Dies* was randomly selected to be held back when fitting the model. Instead, this was used to test the effectiveness of the model at predicting the killer. (Spoiler alert!)

The killer is Jane Wilkinson, the wife of Lord Edgware, and hence also a relative of the victim. Her occupation is an actress, which places her in the "Careers" category. According to the model, Jane Wilkinson has the second largest odds of being the murderer after Geraldine Marsh, a child, and Christie surely wouldn't... would she? The results, though, highlight one of the limitations of the model - its inability to really distinguish between the characters.

Covariate	**Odds Ratio**	
	Estimate	Confidence Interval
Relative	2.13	(1.17, 3.80)
Occupation		
Servants, e.g. butlers, secretaries	1.82	(0.88, 3.80)
Public services, e.g. doctors, clerics	1.03	(0.46, 2.25)
Careers, e.g. solicitors, authors	1.19	(0.54, 2.56)
Unpaid roles, e.g. aristocrats, housewives	1.38	(0.70, 2.70)

Table 1: The odds ratios from fitting the logistic regression model of whether a character is a murderer against their occupation and whether they are a relative. The odds ratios are relative to those with an unknown occupation. The confidence intervals are based on the profile likelihood.

If anything, this study shows why Agatha Christie continues to deceive readers. Yes, relatives are more likely than non-relatives. Perhaps the police can be trusted and the aristocrats treated with suspicion. But more startling is the absence of many associations - such as gender - which suggests no underlying pattern or prejudice. The advice is: leave the detection to Poirot and his "little grey cells"!

Acknowledgements
Thank you to Professor Brian Francis and The Editor for their help with this article.

References
[1] M. Edwards. *The Golden Age of Murder*. HarperCollins, 2015.
[2] V. Ward. 'Experts devise formula to crack Agatha Christie's murder mysteries'. The Daily Telegraph, 2015.
https://www.telegraph.co.uk/news/newstopics/howaboutthat/11779272/Experts-devise-formula-to-crack-Agatha-Christies-murder-mysteries.html