



Thesis MA Cultural Economics and Entrepreneurship  
Faculty of History, Culture and Communication  
June, 2015

Title: Book awards as indicators of literary quality: A quantitative investigation of the level of consensus between expert juries, between expert juries and consumers, and between experts over time

Author: Mariëlle de Jong

Student-id: 411569

Supervisor: Dr. E. Dekker

Second reader: Dr. C. W. Handke

Date: 08-06-2015

## Contents

List of tables	3
Abstract	4
1. Introduction	4
2. Awards as a form of quality assessment in the arts	6
3. Conceptualizing and measuring consensus	10
4. Method	13
<i>Measures</i>	
<i>Sample</i>	
5. Results	15
6. Conclusion	21
Acknowledgement	23
References	23
Appendices	26

**List of tables**

Table		Page
1	Amount of books that win a majority of awards/nominations	16
2	Original results found by Dekker and Popik on the amount of movies that win a majority of nominations	17
3	$\beta$ -values, winners and nominations	18
4	Various indicators reflecting the overlap between the award samples and anthologies	21
5	Characteristics of six book award-granting juries and their prizes	26
6	Overview of the number of books awarded/nominated by all six organizations	27

# **Book awards as indicators of literary quality: A quantitative investigation of the level of consensus between expert juries, between expert juries and consumers, and between experts over time**

## **Abstract**

This study reports three distinct analyses of consensus on literary quality: consensus between expert juries, between expert juries and consumers (non-experts), and between experts over time. The study looks into the verdicts of three American juries (Pulitzer Prize for Fiction, National Book Award for Fiction and the PEN/Faulkner Award for Fiction) and three Dutch/Flemish juries (AKO Literatuurprijs, Libris Literatuurprijs and the Gouden Boekenuil) that annually award books and investigates whether they nominate/award the same books in a given year. This research recognizes the argument raised by Dekker and Popik (2014) which states that this type of interjury consensus cannot be appropriately measured by the means of reliability analysis and the Cronbach's alpha composite. Following Dekker and Popik, different aspects of consensus are discussed, after which the measure ( $\beta$ ) is used to assess the level of agreement. Deviations from the overall consensus by particular juries are also indicated. This is followed by an examination of the accordance between the juries' judgments and consumer preference. The latter is indicated by bestseller lists on the one hand and online ratings on the other hand. Finally, this research employs a test-of-time analysis. It examines whether books that are honoured by the expert juries appear in anthologies too. The analyses reveal an interjury tendency towards dissensus rather than consensus, which contrasts with findings for the film industry reporting moderate to high levels of agreement. Theoretical reasons for this inter-industry difference are put forward. Moreover, results on the level of consensus between expert juries and consumers suggest that consumers also find that the nominated/awarded books are superior to general books. This does however not persuade them to buy these honoured books in great numbers. A final finding is that experts over time agree moderately with each other on literary quality

**Keywords** Cultural economics • Awards • Expert juries • Literary prize • Interjury consensus • Product quality • Book sales • Test of time analysis

## **1. Introduction**

In the past, art evaluations were mostly implicit (Ginsburgh & Weyers, 2014: 291) whereas nowadays the number of awards and other explicit appraisals can hardly be tracked (English, 2005: 17). As a result, the amount of available quantitative data on, inter alia,

literary quality has increased significantly. Now that this information is easily available through the Internet, it is no surprise that researchers have become interested in the relationships between various indicators of artistic quality, such as the relationship between expert judgment and sales figures (e.g. Ponzo & Scoppa, 2015), and the relationship between expert judgment and long-term recognition (Ginsburgh, 2003). An important step to be made before such relationships can be meaningfully interpreted is to establish to what extent experts actually agree on quality: to what extent there is a consensus between various award juries on which artists should be honoured for their outstanding creative achievement. In this thesis, the level of agreement on literary quality between various book award juries was assessed, after which the degree of accordance between these award juries and consumers (i.e. non-experts) and the level of consensus between literary experts over time was examined.

The research on award-granting juries is interesting for various reasons. First of all, there is no cultural phenomenon wherein art's relationship to money, to politics, to the social and the temporal is given greater emphasis. In fact, research on expert juries involves questions of power, of what constitutes specifically cultural power, and how this power is obtained. It involves the question of cultural prestige, how is such prestige produced, and what rules govern its circulation (English, 2005: 3). What is more is that the cultural award touches every corner of the cultural universe nowadays (cinema, television, music, literature, painting, architecture, history, dancing, journalism, photo-graphy, comedy, fashion, etc.) while at the same time very little is known about these spectacles. Some have even claimed that “there is no form of cultural capital so ubiquitous, so powerful, so widely talked about, and yet so little explored by scholars as the cultural prize” (English, 2002, p. 109). Moreover, it is of interest to know whether consumers looking for a guide telling them what is worthy of their time and money can rely on the judgments of award juries or whether awards are indeed the mere machinery of cultural production – as is often asserted. Not to mention that the weeks before and after the ceremonies of famous award series are important periods for the entire industry. It would be interesting to know how the industry can capitalize on these events in terms of sales or educational activities. Lastly, there has been much ado about juries nominating/awarding books on the basis of matters other than quality, but also the limited number of female nominees/winners and jurors has been a point of discussion, yet whether these criticisms are justified is unknown.

From a cultural economic point of view, research on the accordance of expert opinion contributes to the underlying debate on the convergence or divergence of taste of individuals (Dekker and Popik, 2014); a debate that more than 30 years after the publication of the famous essay ‘De

*Gustibus Non Est Disputandum*<sup>1</sup> by neo-classical economists Stigler and Becker (1977) has still not resulted in some genuine theories other than those conveniently assuming that individuals have identical taste (Towse, 2010; Blaug, 2001). This is especially relevant since experts are said to provide one of the most important signals about the quality of products in the creative industries which are defined by great uncertainty and risk (Wijnberg, 2011; Caves, 2000).

The study at hand sets out to indicate the degree of three types of consensus on literary quality. First, the degree of consensus between award-granting juries selecting the best literary works out of the same group of contesting books was looked at. Second, the level of agreement between expert judgment and consumer preference was examined by looking at bestseller lists on the one hand, and online ratings on the other hand. Finally, expert consensus over time was estimated by examining whether nominated/awarded books also appear in anthologies that list critically acclaimed or historically significant books. It is one of the few studies in the literature on cultural awards to focus on interpanel consensus and the first attempt to do so for the book industry. Besides, it is the first paper that examines several types of consensus instead of focusing on only one type. As such, this study makes a unique contribution to the cultural economic literature that sets out to define the role of critics in the creative industries. The article starts with an introduction on quality assessment in the arts, followed by a conceptualization of interjury consensus and a discussion on the methods, after which the results and conclusions are presented.

## **2. Awards as a form of quality assessment in the arts**

The aesthetic evaluation of artworks - including literature - is, and always has been, a very controversial exercise (Ginsburgh & Weyers, 2005: 1). On the one end of the continuum there are those who support the view that artistic quality lies in the artwork itself (see, for instance, De Piles, 1708), while on the other end there are those who like Hume (1757) believe that “beauty is no quality in things themselves: it exists merely in the mind which contemplates them; and each mind perceives a different beauty” (p. 6).

Although the divergence of these philosophical views is already interesting in its own, it would be a rather unfamiliar subject to most if only the market was not involved in the exchange of cultural products. However, the conflict between culture and the market is as old as the hills and anyone ever having to buy a ticket for a new play or having to choose a new novel to read knows that difficult choices have to be made: choices that are not only based on price but also on perceived

---

<sup>1</sup> The neo-classical economists Stigler and Becker famously argued that “tastes neither change capriciously nor differ importantly between people. [Tastes] will be there next year, too, and are the same to all men” (p. 76). Accordingly to Blaug (2001) “it has received some of its best criticisms from cultural economists” (p. 125).

quality, which fundamentally makes the aesthetic evaluation of artworks a major challenge from an economical perspective (Karpik, 2010: 3).

Ever since its birth, cultural economics has attempted to bridge this gap between what Klammer (2014) calls: the financial conversation and the cultural conversation. Among other things, it has made numerous efforts to measure artistic quality and the perceived source of quality (artwork versus mind) has led scholars to use different approaches. A first distinction is in qualitative and quantitative assessment methods. The qualitative way to assess quality in the arts traditionally comprises descriptions of the artworks with respect to some accepted scheme of values and characteristics (Dekker and Popik, 2014: 95), usually called “properties” by philosophers and “qualities” by economists (Ginsburgh & Weyers, 2005: 1). In literary arts the qualities described by the experts may pertain to choice of words, syntactic complexity, narrative structure, originality, etc.

In addition, there are three quantitative methods to explain the differences in quality between artistic products (Ginsburgh & Weyers, 2005). The usual method was (and still is) to decompose the product into fundamental characteristics and rate these on a certain scale to determine overall quality. A distinction must be made here. Some researchers suggest that the relative value of characteristics is determined by the perfect competitive market (hence objective) and remains the same whatever the goods in which this characteristic is embodied (Lancaster, 1966). Others assert that these relative values are subjective. Karpik (2010) and others have thoroughly argued that the first argument is not valid for cultural products, because these products are unlike other differentiated products (such as houses) not the *aggregate* of their characteristics, but just the *structure* of their characteristics. This is well illustrated by the costs incurred for the restauration of the painting “Who’s Afraid of Red, Yellow and Blue III” by Barnett Newman. Over one million guilders ( $\pm$  half a million euros) were spend after a man damaged the painting (Fontein, 2013). What is clear is that the theory based on the premise that products are a “bundle of characteristics” is not able to explain why a product that mainly consists of canvas and paint would be deemed so valuable. Still, there are many art critics, philosophers, economists and others who claim that there are general standards which make a work “good”, as opposed to those who sustain that there exist no such general standards, and that every quality is contextual (Ginsburgh & Weyers, 2005: 2).

Short of being able to decompose artworks into characteristics and rating them, an alternative method of ranking is to see which works and artists survive over time and space (Ginsburgh, 2003: 100). The method known as the “test-of-time” revolves around the idea that the passing of time makes it possible to reduce some of the noise present in evaluations made shortly after the work is produced like fads, clearing the way for transcending works. Besides, the “test of space” is based on the idea that the more wide an artwork is spread, the more sincere the admiration is. Until

recently, these methods were hardly used by (cultural) economists (Ginsburgh & Weyers, 2005: 5), yet some contributions are now available. For instance, Vermeulen, Van Dijck and De Laet (2013) investigate whether leading Dutch and Flemish painters from the 17th century onwards endured the are nowadays still known because of historical processes or because they were recognized by art lovers and critics in the course of art history.

The last quality assessment method puts the burden of proving quality on judges. Traditionally, the judges comprise of experts, although economists often argue that the actual choices made by consumers are a better measure (Ginsburgh, 2003: 100). As Dekker and Popik (2014) point out, this method is based on counting various success indicators like the peer or industry acclaim (expressed in prizes and awards), critics' opinions (expressed in the amount of either stars awarded or publications that mention a title); and sales figures and other earnings (hence, consumers approval).

Expert judgments in the form of prizes and awards date back to at least the sixth century B.C.E. when contests or competitions combining music, poetry, and drama were routinely organized throughout the cities of east central Greece (English, 2005: 30). Countless prizes have been in existence ever since and some of them, like the Prix de Rome and the Nobel Prize for Literature, have a long ongoing history. The great impact that prizes may is not a new phenomenon either. Rousseau, for example, only became a famous philosopher after he wrote the winning response to an essay contest conducted by the Academy of Dijon in 1750 (Delaney, n.d.). Acknowledging these facts, English (2005), author of the book *The Economy of Prestige: Prizes, Awards and the Circulation of Cultural Value*, argues that the last century has demonstrated a "tremendous growth of cultural prizes, which have been expanding in number and in economic value much faster than the cultural economy in general" (p. 10). Although one could argue that this statement is hardly verifiable for the very long-term, English convincingly shows that the ratio of literary prizes to new titles has risen tenfold between 1920 and 2000. He also found that by the end of the twentieth century there were more film awards distributed every year than full-length films being produced. This leads to the seemingly straightforward question: how is it that awards are so popular?

Numerous reasons from various disciplines have been put forward. Economists maintain that the popularity of the cultural award is the result from the fact that cultural products are experience goods for which (the majority of consumers) are unable to evaluate the quality before or even after consumption (Caves, 2000). Consequently, rational consumers look for surrogate signs of quality, which can come from a variety of sources including advertising (Akerberg, 2003), learning from peers (Beck, 2007), branding (Montgomery & Wernerfelt, 1992), and indirect signals from firms including price, quantity or advertising decisions (Caves & Greene, 1996). For cultural products, expert judgments are assumed to form an important source of information on quality (Wijnberg, 2011: 63).



Awards and prizes have been criticized from all different directions. Some disapprove their way of doing things and others simply grumble by a conception of art as a contest of competition. Nevertheless, there can be no denying that experts and other gate-keepers play an important role in the functioning of the cultural sector of an economy. Be that as it may, economists and other scholars have traditionally paid little attention to the phenomenon in itself (Cameron, 1995: 321). A current boom in cultural economics however is in studies on the relationship between awards and other success indicators. In particular the relationship between awards and box-office success has been a subject of debate and has been studied in, *inter alia*, the context of competitions for classical music (Ginsburgh & van Ours, 2003), Broadway theatre (Reddy, Swaminathan & Motley, 1998), the music record industry (e.g. Watson & Anand, 2006), wine (e. g. Friberg & Gronqvist, 2012), and the motion picture industry (e.g. Deuchert, Adjamah & Pauly, 2005). Evidence on the commercial effect of book awards is relatively scarce, which fits the image that is portrayed by Blaug (2001) and Towse (2010) that states that the whole field of literature takes a back seat within cultural economics. However, some contributions are available (Ginsburgh, 2003; Clement, Proppe and Rott, 2007; Ashworth, Heyndels and Werck, 2010; and more recently Ponzo and Scoppa, 2015). It is surprising to note that only one study to date has compared the differences in the commercial impact of multiple awards (Gemser, Leenders & Wijnberg, 2008). Overall, the results of studies on the predictive value of cultural prizes in terms of box-office success vary widely –even for studies within the same creative industry.

Relationships between awards and success indicators other than sales figures have also been researched, albeit infrequently. For instance, Ginsburgh (2003) puts the judgement of juries to the test by researching whether expert opinions can be considered a good indicator of the true aesthetic quality of an artistic product. By conducting three test of time analyses, his results suggest that this is not the case for both the music, film and book industry, since awards did not prove to predict long-term survival in the market (which Ginsburgh assumes to encapsulate the product's fundamental aesthetic quality). There are, however, some serious issues with this conclusion. In the case of books, Ginsburgh compared fiction books that were just shortlisted for the Booker Prize to books that won the Booker Prize on the basis of the number of editions published 11 to 20 years after the ceremony. Ginsburgh found that in only three times award winners outperformed shortlisted titles, while in nine cases, shortlisted titles did better, and in two other cases winning and shortlisted titles were tied. Based on these findings for only 15 books (!), Ginsburgh (2003) concludes that there is no difference between shortlisted and winning books. Hence, awards cannot be seen as a good indicator of true aesthetic quality. Obviously, his results would have been more convincing if he compared awarded/nominated books to book that were not honoured.

Next to consensus over time, across places, and between experts and consumers, another type of agreement on artistic quality can be distinguished: consensus between expert juries. The succeeding section deals with measurement issues relating to interjury accordancy after a discussion on decision-making procedures within expert juries.

### **3. Conceptualizing and measuring consensus**

In general, consensus is of interest when judges need to score behaviours that cannot be objectively defined in a simple right/wrong sense, but instead require some rating of the degree to which observed behaviours represent particular levels of a construct of interest (Stemler, 2004). Consensus within an award jury is usually reached through discussion. In the case of book awards, the juries (which typically comprise of five to six jurors) first decide which jurors are going to read which books. This is necessary because the amount of submitted books is typically so large that it is impossible for them to read every book. Each book is then read by at least two jurors. Based on their judgment it is decided which books are going to be read by all jurors. Subsequently, the jurors meet on several occasions to choose the longlist nominees, shortlist nominees, and finally the winner. Thus, the announcements of the book award juries are already the outcome of an agreement. Other ways of decision-making may also occur. For instance, the famous Oscars conferred each year by the Academy Awards are based on the votes of hundreds of experts in various domains of cinematic achievement (Simonton, 2004). Many researchers claim that voting is the more powerful way to come to an agreement because it reduces errors (e.g. Larrick & Soll, 2012). What is thought-provoking in light of this claim however is that film award juries appear to be presented with a wider range of criticism than book award juries. Obviously, verdicts in both industries cause newspapers and forums to report that years' inexcusable omissions and puzzling blunders, yet what stands out is that film awards are more associated with criticisms stating that the juries are subject to "behind the scenes" political maneuverings, advertising campaigns, and other arbitrary events (e.g. Holden, 1993). Research is needed to determine whether these issues are related to each other.

In addition to consensus within an expert jury (be it reached through discussion or counting votes), consensus between various expert juries can be distinguished. As Dekker and Popik (2014) point out, it is important to establish to what extent experts actually agree on quality before the relationships between experts and other success indicators can be meaningfully interpreted. To demonstrate this somewhat clearer, think of how meaningful it would be to investigate the relationship between an expert's judgment of a particular book and the long-term recognition of that book, if the expert evaluated this book much too positively according to most of his colleague experts.

By examining the level of consensus between individual award juries one gains knowledge about whether these juries, claiming to select the best artworks in a given year, use the same standard of quality and whether they similarly define creative achievements in a particular creative industry. Research on this type of consensus also clarifies whether juries are in competition or corroborate on a success of a single artwork (Dekker & Popik, 2014). To date, two contributions on the level of agreement between various award juries are available, both focusing on the film industry. No evidence is available for the book industry.

More than anything else, studies on interjury consensus within the creative industries show that the choice for suitable methods is not an obvious one. The first attempt has been by Simonton (2004). He examined a group of seven Anglo-Saxon film award organizations and looked into the rulings of 28 years of ceremonies in ten major award categories. Simonton chose to use a reliability analysis to measure the level of interpanel agreement and found Cronbach's alpha values ranging from .87 to .59. As the literature generally accepts .7 as a threshold, Simonton concluded that the consensus on cinematic quality is relatively high. Furthermore, his results showed that removing juries from the aggregate decisions of the others nearly consistently lowered reported values. Removing the Oscars usually led to the largest drop in the value of alpha. Simonton interprets this in such a way that all juries contribute to the established consensus, and as a sign that the Oscars have an exceptional expertise on the quality of motion pictures.

However, Dekker and Popik (2014) point out that there are fundamental issues with the use of Cronbach's alpha if one is to measure this kind of interjury consensus. This has everything to do with the fact that data on the verdicts of awards juries typically contain only three scoring categories: winning an award, nominated for an award, and no award or nomination. Consequently, alpha values derived from this type of data do not reflect the extent to which juries have given similar ratings (as would be the case if instead a 1-10 or 1-100 scale was used to score the films), but rather reflect the extent to which juries all award or not award the same movies. This is problematic because the mere interest here is in the level of consensus on successful films, while alpha is equally affected by consensus on not nominating a movie.

Dekker and Popik (2014) state that it was not easy to find an alternative method that could discriminate between the two types of consensus identified: consensus on quality and consensus on films that should not be honoured. No measure could tell the whole story. Existing measures like correlation analyses and latent-class analysis did either not match the almost dichotomous data or the research question. Therefore, they were compelled to go back to basics and looked at different dimensions of consensus. First, a small number of thresholds for consensus was set, such as the percentage of winning films who received at least three out of seven awards. This made it possible to compare the level of consensus between the different categories. Results showed, for instance,

that the degree of consensus was higher for the director category than for the other categories. A downside of this method was, however, that there was no absolute standard to compare against. Besides, this method does just like Cronbach's alpha not control for the length of the dataset itself. To tackle this latter problem, Dekker and Popik (2014) constructed a new measure,  $\beta$ , to indicate consensus. With  $pw$  being potential winners,  $aw$  being the actual unique winners, and  $mw$  the minimum potential winners. This formula was also used for nominations, by substituting the relevant variables.

$$\beta = (pw-aw)/(pw-mw) \quad (1)$$

In contrast with Simonton's (2004) findings, results obtained with this method revealed overall values around the .5 mark, indicating the middle ground between consensus and dissensus. As Dekker and Popik (2014) point out, these lower values were not only to be expected on the basis of methodological arguments, but also on the basis of theoretical arguments.

Many factors contribute to what can be hypothesized about the relations between expert juries and their announcements. The most non-political and non-economical argument in favour of consensus is that it is conceivable that people, especially those professionally involved in one of the creative industries, share a certain view on exceptional creative achievement – regardless of the organization they happen to be member of (Dekker & Popik, 2014). This is also sustained by the fact that juries use the same or similar words and phrases to describe an award winning book. “The reader gets sucked into the story”, “surprising” and “magnificent” are repeatedly used and they all signify similar if not the same things. At the same time these phrases remain quite vague: no emphasis is given to a book's topic or its style, therefore it could be expected that similar verdicts will be given out. Another reason to expect a high level of consensus between expert juries is that contagion may occur. As the award season is a cycle of events, experts themselves may trust and follow each other when it comes to naming winners.

It is no secret that the first argument in favour of high levels of consensus is strongly advocated by Simonton (2004). There are however also grounds for low levels of interjury consensus. Dekker and Popik (2014) outline some of them. The most straightforward one is that the great number of awards in itself creates divergence. If juries were all just confirming each other's decisions, one award series would be enough. In other words, it is divergence that provides juries their *raison d'être*. Another more economic reason for divergence mentioned by Dekker and Popik is that award ceremonies are important events for an entire creative industry where famous individuals attract additional attention. That attention then can influence earnings for the nominees and winners and

their future projects, but also for the entire industry. In this regard, it would be disadvantageous to honour the same artists over and over again.

An additional anti-contagion effect that should be mentioned here is that many award organizations aim to promote cultural practice within a particular creative industry. Not uncommonly, their recognition is accompanied by a sizeable donation. Awarding the same artist twice would then undermine the organization's mission. What stands out with respect to this research, is that Dutch/Flemish literature award organizations grant far larger sums of money than American literature award organizations, yet the American award series are much more famous. For example, the winner of the Libris Literatuurprijs receives €65.000 where the winner of the Pulitzer Prize for Fiction receives the equivalent of €8981.50 (see appendix 1).

#### **4. Method**

Although the objective of this study was to gain knowledge about expert juries, rather than about the cultural products themselves, the data that needed to be gathered are book titles – the winners, and nominees honoured annually by different organizations. Consequently, the data came primarily from electronic sources, like various official sites, such as those for the AKO Literatuurprijs (<http://www.akoliteratuurprijs.nl/>), the Pulitzer Prize for Fiction (<http://www.pulitzer.org/>), and the online book database Goodreads (<https://www.goodreads.com/>). Where needed, information was supplemented with information from other sources such as online newspaper articles. Where possible, information was cross-checked using databases provided by award organizations. These cross-checks revealed that there were no discrepancies in the assignments of the awards across alternative sources.

#### *Measures*

Measures were defined for all organizations. The following ordinal scheme was used to define the scores: 2 = recipient of an award, 1 = received a nomination, and 0 = received neither an award nor a nomination. For the Dutch/Flemish juries, longlist nominations were considered in addition to shortlist nominations. Analyses that include these announcements were based on a different scheme: 3 = recipient of an award, 2 = received a shortlist nomination, 1 = received a longlist nomination, and 0 = received neither an award nor a nomination.

#### *Sample*

The main sample consisted of all Dutch-language books released in the 1995-2014 period and all American books in the 1981-2014 period that received awards or shortlist nominations. Furthermore, the award or nomination had to come from at least one of the three professional

societies that were taken into account for each country/language area: the Pulitzer Prize for Fiction, the National Book Award for Fiction and the PEN/Faulkner Award for Fiction for America, and the AKO Literatuurprijs, the Libris Literatuurprijs and the Gouden Boekenuil for the Netherlands/Flanders. The American dataset consisted of 384 works and the Dutch-language dataset of 257 works. The Dutch-language sample was extended with longlist nominees from 2007 onwards. This led to an increase of 236 books, so that the total dataset including these announcements consisted of 493 books. An extension with longlist nominee was not possible for the American award series, as not all juries announce these verdicts.

These award juries were chosen for three reasons. First, all organizations have consistently granted annual awards. Second, all awards have been long in existence: at least 20 years for the Dutch/Flemish awards and at least 34 years for the American awards. Third, they all covered the same group of contending books: American or Dutch-language literary fiction for adults published in the year prior to the award ceremony. This is crucial for the measurement of consensus, as looking for an overlap in the announcements would otherwise not make sense. It must be noted however, that some award juries were slightly more open to alternative book types, like non-fiction books and comic books where others were stricter in the eligibility of contesting books. These differences were controlled for during the data-analyses.

For all books, additional data were gathered. First, consumer preference was taken into consideration by consulting annual bestseller lists. Data collected by Publishers Weekly on the ten bestselling novels per year were used for the American dataset, which were available for the whole 1981-2014 period. A major finding was that only 1 out of 293 novels had been a bestseller (*Lila* by Marilynne Robinson). For the Dutch-language announcements, data collected by the Stichting Collectieve Propaganda voor het Nederlandse Boek (CPNB) on the ten bestselling books (so not only novels) per year were used, which were available as of 1999. For the 1995-1998 period no reliable source on the commercial success of books was found. The Dutch-language sample for this period consisted of 448 books, of which 5 were bestsellers at a certain point. A second indicator of consumers' taste that was used are ratings on the world's largest site for readers and book recommendations "Goodreads.com". Goodreads.com is the book version of the website IMDB.com where visitors can express their opinions on a five point scale. Average ratings that were composed of less than ten individual ratings were excluded, since otherwise there would be too much focus on individual opinions rather than on consumer consensus. Average ratings for six books in the American sample were discarded as opposed to 129 ratings for Dutch-language books, showing that the website is more popular in America than in the Netherlands/Flanders. Third, the number of evaluators rating each book on Goodreads.com was noted for all remaining books. *The Poisonwood Bible* by Barbara Kingsolver and *Naar de Overkant van de Nacht* by Jan van Mersbergen were most

often rated (471667 and 4377 times respectively). Fourth, the type of book was recorded with one category for novels and one category for “other books” like biographies, audiobooks, and short stories. This distinction was made to control for the slight differences in the eligibility of books by the different award juries. For the Dutch-language verdicts 376 books were regarded as novels and 117 books were regarded as other books. The American verdicts comprised of 293 and 91 books respectively. Fifth, the nationality of the author of each book in the Dutch-language dataset was noted. This showed that quite a few more Dutch than Flemish authors were awarded or nominated (390 compared to 103). Finally, it was looked whether the books in this study had been included in anthologies. For the American dataset, an online anthology compiled by Daniel Immerwahr (n.d.) on the most “critically acclaimed and historically significant books” of the 20<sup>th</sup> century by year of publication was used. Immerwahr is a scholar and teacher and composed this list by consulting numerous sources, including the Modern Library’s list of the hundred best novels and nonfiction books of the 20<sup>th</sup> century and the chronobiology of historically significant books listed in the back of Hollinger and Capper’s (2006) book *The American Intellectual Tradition*. For the 1981-1999 period, 115 fiction and nonfiction books were recorded. A further examinations revealed that 24 of these books are novels written by American citizens, of which 16 are also listed in my dataset. Unfortunately, no such anthology was available for Dutch-language books. Therefore, it was chosen to use the most relevant anthology available, even though this contribution covered a very recent period. To wit, an anthology compiled by Wim Brands, one of the major advocates of the Dutch literature and host of the television program “Boeken”. He collected sixty fragments out of sixty books written by sixty of the most talented Dutch/Flemish authors of the 21<sup>st</sup> century. Out of the sixty books, twenty were also listed in my sample.

In addition to these main samples, “general samples” were composed which contained fiction books published in the same periods as the books in the “award samples”. This was done so that the rating scores on Goodreads.com of the books in the award sample could be compared with rating scores of novels that, as a rule, did not receive a nomination or an award. The general samples were constructed with the use of online libraries like those of the Koninklijke Bibliotheek (national library of the Netherlands). For every year, the first four novels that appeared in these online libraries were listed.

## 5. Results

This study aims at estimating the level of three types of consensus with respect to book award juries. Three American prizes (Pulitzer Prize for Fiction, National Book Award for Fiction, and PEN/Faulkner Award for Fiction) and three Dutch/Flemish prizes (AKO Literatuurprijs, Libris Literatuurprijs, and Gouden Boekenuil) were considered. Following Dekker and Popik (2014), the

level of interpanel consensus was measured by the use of two basic methods, the most appropriate methods to indicate this type of consensus available to date, and assumingly one of the few methods applicable here. Consensus between experts and consumers on the other hand was measured by means of a comparison between nominated/awarded books and non-honoured books in terms of sales and ratings. This section ends with a test-of-time analyses to see how many books survive over time. By combining these methods, we can critically address the value of award juries as an indicator of literary quality.

Before discussing the findings, it must be noted that some of the results presented here could not be calculated with the use of advanced statistical programmes and instead required quite some manual calculations. For the benefit of transparency, it was chosen to summarize the number of books awarded/nominated by the individual juries in appendix 2. Moreover, all values indicating the level of consensus were, as expected, somewhat higher when only novels were considered instead of all books. The differences in values were, however, so small that it was chosen to just report the values reflecting all books.

#### *Consensus between award juries*

The first question to be addressed is to what extent award juries agree on which books contain high levels of literary quality. Analogous to the study by Dekker and Popik (2014), the first method that was used to measure consensus has been by setting certain thresholds for consensus. Because only three awards were included per country/language-area, it was only meaningful to look at books that won two awards (no books won all three awards), books that were nominated by at least two juries, and books that were nominated by all three juries (nominations include award winners). This information is summarized in Table 1 for all American and Dutch-language verdicts. The percentages were calculated by comparing the amount of individual books winning two awards with the total number of books winning an award, and by comparing the number of books that received two or more nominations with the total number of books. With respect to the Dutch-language verdicts, values are given separately for data that did not include longlist nominations. This was done for the purpose of comparison.

**Table 1.** *Amount of books that win a majority of awards/nominations*

	2 Awards	% of Wins	Nominated by at least 2 organizations	% of Total Amount of Books	Nominated by all three organizations	% of Total Amount of Books
American awards	6	5.08%	54	14.06%	8	2.08%
Dutch/Flemish awards	3	5.26%	58	22.57%	12	4.67%
<i>included longlists</i>	3	5.26%	134	27.18%	36	7.30%

*Note.* PL= Pulitzer Prize for Fiction. N=National Book Award for Fiction. PF=PEN/Faulkner Award for Fiction. A=AKO-literatuurprijs. L=Libris-literatuurprijs. G=Gouden Boekenuil. Nominations include award winners.



These data show that, on average, every six years roughly one American and one Dutch-language book receives two awards (remember that the databases cover 34 years and 20 years respectively). Thus, the degree of consensus on award winning books is similar for the American and Dutch-language verdicts. In contrast, consensus is more prevalent for the Dutch/Flemish nominations than for the American nominations. Data on the Dutch/Flemish verdicts furthermore show that the level of consensus increases when longlist nominations are taken into consideration too.

The results found for award *winning* books cannot be easily compared with those found by Dekker and Popik (2014), since they took seven instead of three juries into account and considered 3 to 4 out of 7 awards reflecting considerable agreement. However, if we roughly compare the results found here with the results found by these researchers it can be stated that interjury consensus on literary quality is quite a bit lower than interjury consensus on cinematic quality.

With regard to the level of agreement on nominations, both Dekker and Popik (2014) and this research took three organizations into account (information on nominations was not available for the four remaining film award organizations), so that results on this aspect of consensus could be compared directly. This comparison revealed an even greater difference in agreement levels. In fact, results found by Dekker and Popik demonstrate that, on average, 46.03% of films included in each category is nominated by two out of three organizations (see Table 2). This contrasts sharply with the 14.06% of American books and the 22.57% of Dutch-language books that win a majority of nominations. The same difference becomes apparent when we look at the three out of three threshold: an average of 19.66% of films were nominated by all film award juries, compared to only 2.08% (American juries) or 4.67% (Dutch/Flemish juries) of books.

**Table 2.** *Original results found by Dekker and Popik on the amount of movies that win a majority of nominations*

	2+	% of	3	% of
	<i>nominations</i>	<i>nominations</i>	<i>nominations</i>	<i>nominations</i>
Picture	150	48.39%	64	19.94%
Director	113	57.07%	59	27.57%
Male lead	133	10.92%	56	16.47%
Female lead	132	42.17%	56	17.23%
Male support	111	48.26%	34	13.33%
Female support	104	46.64%	36	13.90%
Cinematography*	56	31.82%	-	-
Screenplay	171	50.00%	74	20.96%
Foreign language	144	48.98%	82	27.89%
Overall	123.7	46.03%	57.6	19.66%

The second method employed to measure consensus has been by calculating  $\beta$ -values as described in chapter 3. Table 3 provides the results found with this measure. Two issues were encountered. First, the National Book Award jury awarded three books more than the expected 34 and the Pulitzer Prize for Fiction awarded one book less. This was because the National Book Award disclosed

separate prizes for hardcover and paperback fiction books from 1981-1983, which were all included in the dataset, and because the Pulitzer Prize for Fiction did not chose a winner in 2014. Second, some juries nominated considerably more books in a given year than other juries did (see appendix 1), thus significantly increasing the *mw* value for the calculations on nominations. For instance, the Libris Literatuurprijs nominated 100 books for their shortlist in the 1995-2014 period while the Gouden Boekenuil only nominated 68 books. To overcome these minor issues, the average number of winning and nominated books were used instead of the maximum amount of awards/nominations by a particular jury. This did not affect the scores greatly.

**Table 3.**  $\beta$ -values, winners and nominations (nominations include winners)

	American awards				Dutch-language awards			
	$\beta$	PL	N	PF	$\beta$	A	L	G
Winners	.10	.06	.09	.09	.08	.05	<b>.10</b>	.00
Shortlist nominations	.21	.12	.17	.20	.32	.27	.30	.19
Longlist nominations	-	-	-	-	.38	.31	.31	.32

*Note.* Bold values indicate higher (or equal) than overall scores. PL= Pulitzer Prize for Fiction. N=National Book Award for Fiction. PF=PEN/Faulkner Award for Fiction. A=AKO-Literatuurprijs. L=Libris-Literatuurprijs. G=Gouden Boekenuil.

In accordance with the results found with the first method, the results in Table 3 show that consensus on award winning books is highly similar for American and Dutch-language juries and that consensus on shortlist nominations is greater for the Dutch-language verdicts than for the American verdicts. Moreover, consensus is greatest on longlist nominations.

The overall values for the  $\beta$  measure for winners as well as for nominations are remarkably low. For awards they are around the .1 mark and for nominations around the .2 (American juries) or .3 mark (Dutch/Flemish juries), indicating a strong tendency to dissensus rather than consensus. In contrast, overall scores found by Dekker and Popik (2014) were around the .5 mark for award winning films and around the .6 mark for nominated films. Again, this suggest that interjury agreement on literary quality is much lower than interjury consensus on cinematographic quality.

#### *Significant differences between juries*

Next to the examination of consensus between juries, deviations from that consensus by particular juries were also looked at. Table 3 shows the changes in  $\beta$ -values when a particular jury is omitted from the analysis. What can be seen is that omitting one of the juries from the composite usually led to a drop in consensus. For instance, when the National Book Award for Fiction was deleted from the nominations composite,  $\beta$ -values dropped from .21 to .17. Two points should be concluded from these results. First, the decrements to the consensus were largest for the Pulitzer Prize for Fiction in case of the American verdicts and the Gouden Boekenuil in case of the Dutch-language verdicts (omitting these from the composite resulted in the greatest drops). This was supported by data that show that these juries award or nominate less books that receive no other awards or

nominations than the other juries did. Besides, in only one occasion did the deletion of a jury actually increase  $\beta$ . This happened when the Libris Literatuurprijs was omitted from the score measure corresponding to the level of consensus on award winners. This seems to point out that the Libris Literatuurprijs betrayed assessments that go against the implicit consensus witnessed in the remaining awards. However, since the level of consensus is generally so low, these results are very sensitive to particular cases. For instance, if the Libris Literatuurprijs awarded one more book that was also awarded by one of the other two Dutch/Flemish juries, consensus would drop from .10 to .05.

#### *Consensus between expert juries and non-experts/consumers*

The second issue that was addressed is whether expert juries and consumers agree on quality. Two types of data were used as indicators of consumer preference. First, bestseller lists were consulted to determine which books were among the ten most financially successful in a given year. As was stated earlier, very few books appeared both in bestseller lists and this study's datasets. In fact, only 1 out of the 384 fiction books that were included in the American dataset had been a bestseller compared to 5 out of 390 Dutch books (Flemish books are not included here because data collected by the CPNB only concern the Netherlands). What must be noted is that these results cannot be directly compared, because the American list reported only bestselling fiction books whereas the Dutch-language list reported both fiction and non-fiction books. Despite the fact that there are not data to compare against, it can be stated that the level of agreement between expert juries and consumers in terms of sales is somewhat higher in the Netherlands than in America, yet still very low.

Besides, the accordance between expert juries' taste and consumers' taste was examined by comparing the average rankings for awarded/nominated books to general fiction books published in the same time period. Three independent samples t-tests were conducted: one for the American announcements, one for the Dutch-language announcements, and one for the Dutch-language announcements including longlist nominations. In the American case, results revealed a significant difference in scores between books in the general sample ( $M = 3.23$ ,  $SD = .47$ ) and books in the award sample ( $M = 3.73$ ,  $SD = .28$ );  $t(445) = 11.994$ ,  $p < .001$ . The effect size for this analysis ( $d = 1.29$ ) was found to exceed Cohen's (1988) convention for a large effect ( $d = .80$ ). The t-test conducted for the Dutch/Flemish verdicts also revealed a significant difference in scores between books in the general sample ( $M = 3.14$ ,  $SD = .48$ ) and books in the award sample ( $M = 3.43$ ,  $SD = .36$ );  $t(266) = 5.142$ ,  $p < .001$ ,  $d = .68$ . Results for the Dutch/Flemish verdicts were similar when longlists were taken into consideration;  $t(432) =$

5.373,  $p < .001$ ,  $d = .64$ . From this we see that American and Dutch-language books that are honoured with nominations and/or awards are more appreciated by consumers than books that are typically not honoured by expert juries. This effect is especially strong for American works of fiction.

Another three independent samples t-tests were conducted to see whether books that win at least one award receive even higher scores than books that do not win an award but instead only received (a) nomination(s). Regarding the American verdicts, there was a significant difference in scores between winning books ( $M = 3.79$ ,  $SD = .23$ ) and non-winning books ( $M = 3.70$ ,  $SD = .29$ );  $t(376) = -2.990$ ,  $p < .01$ ,  $d = .34$ , with winning books receiving somewhat higher average rankings than non-winning books. No significant differences between winning books and non-winning books were found for the Dutch-language announcements.

These two-sided data on the accordance between expert judgment and consumers' taste appear to suggest that although consumers appreciate awarded/nominated books more than general works of fiction, these books are not bought at great numbers. This puts forward the question whether the books that consumers actually buy are then books they appreciate even more, or whether perceived quality is not such an important determinant of consumer choice after all. To answer this question, however, additional data needed to be gathered, so that average rankings as displayed on Goodreads.com were consulted again. This time, rankings of around 70 bestselling books for each dataset were collected and compared to the rankings for the awarded/nominated books. An independent samples t-test showed that American bestselling books ( $M = 3.87$ ,  $SD = .21$ ) received higher scores on Goodreads.com than awarded/nominated books ( $M = 3.72$ ,  $SD = .28$ );  $t(456) = -4.17$ ,  $p < .001$ ,  $d = -.57$ . Similar results were found when Dutch bestselling books ( $M = 3.69$ ,  $SD = .44$ ) were compared to awarded/nominated Dutch books ( $M = 3.42$ ,  $SD = .39$ );  $t(447) = -5.66$ ,  $p < .001$ ,  $d = -.65$ .

#### *Expert judgment and long-term survival*

The final issue that was addressed is whether experts over time agree with each other on which books show exceptional literary quality. The data that were gathered to indicate this type of consensus are summarized in Table 4. Because the anthologies did not cover the exact same periods as this study's datasets, it was first sorted out how many books were awarded/nominated within the period that the anthologies actually covered. Furthermore, it was noted how many books were listed in the anthology and how many of these co-appeared in the datasets. Some basic calculations revealed that there was an overlap of 66.7% for the American anthology and 23.3% for the Dutch/Flemish anthology (or 33.3% when longlist nominations are also included). In many cases

however, these books were recognized by just one out of three award juries; a maximum of 50% of books were recognized by two or all organizations. The overlap between the juries' verdicts and the anthologies were similar for all juries, yet the Pulitzer Prize for Fiction recognized somewhat more of the critically acclaimed or historically significant books that the other American juries did. A major downside to this test-of-time analyses is that there is no data to compare against.

**Table 4.** *Various indicators reflecting the overlap between the award samples and anthologies*

	<i>Period that anthology covers</i>	<i>Number of books in dataset for this period</i>	<i>Number of books in anthology</i>	<i>Number of books that appeared both in anthology and dataset</i>	<i>% overlap between anthology and samples</i>	<i>Number of co- appearing books that were nom./ awarded by at least 2/3 organizations</i>
American verdicts	1981-1999	206	24	16	66.7%	8 (50%)
Dutch-language verdicts	2000-2014	234	60	14	23.3%	5 (36%)
<i>included longlist</i>	2000-2014	435	60	20	33.3%	9 (45%)

## 6. Conclusion

The main focus of this paper was to analyse the value of book awards as indicators of literary quality. Analyses of three types of consensus regarding prominent literature expert juries were conducted in light of this aim. The first question that was addressed was: do expert juries agree on literary quality? Based on the multidimensional methods that were employed, the answer should be 'on the contrary'. In this study, three dimensions of interjury consensus were identified: consensus on particular high-quality books (investigated with the threshold method), the degree of consensus on nominations and being part of the dominant opinion (both investigated using the measure  $\beta$ ). The results found using these measures demonstrate a strong tendency towards dissensus rather than consensus. This contrasts with previous findings for the film industry that revealed moderate levels of agreement between expert juries (Dekker and Popik, 2014).

The causes underlying this inter-industry difference were not investigated, but some things can be mentioned here. An obvious cause would be that the amount of new titles annually released to the public is much greater for the book industry than for the film industry. Accepting that tastes are not identical, this in itself brings about lower levels of agreement. Furthermore, it is conceivable that certain differences in the supply chain of books versus that of films are key. Compared to the film industry, the book market is relatively simple. Unlike movies, books do not involve a complex set of different professionals to interact. Where movies may be killed by many parts of the chain, for instance when the different professionals do not get along, books only have to endure the judgment of a single publisher (Canoy, van Ours & van der Ploeg, 2006). Moreover, books have much lower production costs so that financing is less problematic. As a result, books face a milder selection at the gate, making after-release interjury consensus less likely.

Secondly, it was found that book award juries and consumers strongly disagree with each other on literary quality, at least, if we assume that sales figures (and bestseller lists in specific) are a good indicator of consumers' taste. The extra analysis that was conducted suggests that this is indeed the case, since bestselling books proved to be appreciated best by consumers, only to be followed by awarded books and ultimately by general books. When we interpreted the findings on the level of consensus between experts and consumers preference as expressed in ratings on the one hand and sales on the other hand altogether, it should be concluded that consumers agree with the experts that awarded books are of a higher quality than general books, but there is another segment of books appearing in bestseller lists that consumers appreciate even more. These findings lend support to the claim that the market for fiction books should not be seen as a single entity but instead consists of several distinct markets. Awarded books can be seen as the equivalent of art-house films and bestselling books as the equivalent of mainstream films. Likewise, awarded books are bought by different kind of people which was already pointed out by Bourdieu (1984) who found that people with a lower level of education hardly buy prize winning books, even when the low amount of books they generally buy is taken into account. Instead, highly educated people rely more on the judgments of award juries and tend to buy these books. Therefore, there seems to be a compelling reason to argue that the bestseller lists used in this research have been a mismatch with the data. Presumably, the use of a bestseller list registering more "highbrow" books would have resulted in higher levels of agreement.

Lastly, the test of time analyses showed that around half of the books listed in anthologies as the most historically significant or critically acclaimed books within a certain time period are picked up by (one of) the juries. At first sight this does not seem to point in the direction of a high level of consensus between experts over time. This can however not be ascertained as no data was available to compare against.

Taken together, these results on various types of consensus on literary quality suggest that there is some information transmitted through awards, especially for those not trusting the alleged wisdom of the crowd. Since books are experience goods and consumers have a hard time deciding which book to buy when left to themselves, some information is better than no information.

As in any study within the social sciences, we must rest on constructions that are never perfect and this has certainly been the case for the present study. Nevertheless, there is no need to say that the current results challenge the idea that literature expert juries generally disagree on which books should be honoured for outstanding creative achievement and, non the least, call for the need to continue research on the convergence or divergence of literature experts' taste. Further research could examine the relative strength of contagion and anti-contagion effects. With regard to consensus between literature expert juries and consumers and consensus between experts over time,

there is clearly considerable scope for future theoretical and empirical development. These levels of consensus are best compared with levels of consensus in other time periods or other industries.

## Acknowledgement

I am grateful to my supervisor Dr. E. Dekker for his support and contagious optimism.

## References

- Akerberg, D. A. (2003). Advertising, learning, and consumer choice in experience goods markets: An empirical examination. *International Economic Review*, 44(3), 1007-1040.
- Ashworth, J., Heyndels, B. & Werck, K. (2010). Expert judgements and the demand for novels in Flanders. *Journal of Cultural Economics*, 34, 197-218.
- Beck, J. (2007). The sales effect of word of mouth: a model for creative goods and estimates for novels. *Journal of Cultural Economics*, 31, 5-23.
- Blaug, M. (2001). Where are we now on cultural economics? *Journal of Economic Surveys*, 15, 123-143.
- Bourdieu, P. (1984). *Distinction: A social critique of the judgement of taste*. Harvard University Press.
- Cameron, S. (1995). On the role of critics in the culture industry. *Journal of Cultural Economics*, 19, 321-331.
- Canoy, M., van Ours, J. C. & van der Ploeg, F. (2006). The economics of books. In V. A. Ginsburgh & D. Throsby (Eds.), *Handbook of the Economics of Art and Culture* (pp. 721-761). Amsterdam: Elsevier.
- Caves, R. (2000). *Creative industries: Contracts between art and commerce*. Cambridge, MA: Harvard University Press.
- Caves, R. E., & Greene, D. P. (1996). Brands' quality levels, prices, and advertising outlays: Empirical evidence on signals and information costs. *International Journal of Industrial Organization*, 14(1), 29-52.
- Clement, M., Proppe, D., & Rott, A. (2007). Do critics make bestsellers? Opinion leaders and the success of books. *Journal of Media Economics*, 20(2), 77-105.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.)*. Hillsdale, NJ: Erlbaum.
- Dekker, E. & Popik, Z. (2014). By unanimous decision? A second look at consensus in the film industry. *Creativity Research Journal*, 26(1), 95-105.
- Delaney, J. J. (n.d.). *Jean-Jacques Rousseau (1712—1778)*. Retrieved from: <http://www.iep.utm.edu/rousseau/>.
- De Piles, R. (1708). *Cours de peinture par principes*. Paris: Gallimard.
- Deuchert, E., Adjamah, K., & Pauly. (2005) For Oscar glory or Oscar money? Academy awards and movie success. *Journal of Cultural Economics*, 29(3), 159-176.
- English, J.F. (2002). Winning the Culture Game: Prizes, Awards, and the Rules of Art. *New Literary History*, 33(1), 109-135.
- English, J. F. (2005). *The economy of prestige: prizes, awards and the circulation of cultural value*. Harvard: Harvard University Press.
- Fontein, J. (2013). Reconstructie: het bizarre verhaal van Who's Afraid of Red, Yellow and Blue III. Retrieved from: <http://www.volkskrant.nl/recensies/reconstructie-het-bizarre-verhaal-van->

- whos-afraid-of-red-yellow-and-blue-iii~a3512580/
- Friberg, R., & Gronqvist, E. (2012). Do expert reviews affect the demand for wine? *American Economic Journal: Applied Economics*, 4(1), 193-211.
- Gemser, G., Leenders, M. A. A. M., Wijnberg, N. M. (2008). Why some awards are more effective signals of quality than others: a study of movie awards. *Journal of Management*, 34, 25-54.
- Ginsburgh, V. (2003). Awards, success and aesthetic quality in the arts. *Journal of Economic Perspectives*, 17, 99-111.
- Ginsburgh, V., van Ours, J. C. (2003). Expert opinion and compensation: Evidence from a musical competition. *The American Economic Review*, 93, 289-296.
- Ginsburgh, V., Weyers, S. (2005). Quantitative approaches to valuation in the arts, with an application to movies. In: M. Hutter & D. Throsby (Eds.) *Beyond Price. Value in Culture, Economics and the Arts* (pp. 179-199). New York: Cambridge University Press.
- Goodwin, C. (2005). Art and culture in the history of economic thought. In V. Ginsburgh & D. Throsby (Eds.), *The handbook of the Economics of Art and Culture*. Amsterdam, Elsevier.
- Holden, A. (1993). *Behind the Oscar: The secret history of the Academy Awards*. New York: Simon & Schuster.
- Hollinger, D. A. & Capper, C. (2006). *The American intellectual tradition, Vol. II: 1865 to the present*. Oxford: Oxford University Press.
- Hume, D. (1757). Of the standard of taste, in *On the Standard of Taste and Other Essays*. Indianapolis: The Bobbs-Merrill Company (1965).
- Immerwahr, D. (n.d.). *The Books of the Century: 1900-1999*. Retrieved from: <http://www.booksofthecentury.com/>
- Karpik, L. (2010). *Valuing the unique: The economics of singularities*. Princeton: Princeton University Press.
- Klamer, A. (2014). *Doing the right thing*. Unpublished manuscript, Department of Cultural Economics, Erasmus University Rotterdam, Rotterdam, The Netherlands.
- Lancaster, K. J. (1966). A new approach to consumer theory. *Journal of Political Economy*, 74(2), 132-157.
- Larrick, R., & Soll, J. (2012). Intuitions about combining opinions: Misappreciation of the averaging principle. *Management Science*, 52(1), 111-127.
- Montgomery, C. A., & Wernerfelt, B. (1992). Risk reduction and umbrella branding. *Journal of Business*, 65, 31-50.
- Ponzo, M. & Scoppa, V. (2015). Experts' awards and economic success: evidence from an Italian literary prize. *Journal of Cultural Economics*. Retrieved from: <ftp://ftp.iza.org/RePEc/Discussionpaper/dp8765.pdf>
- Reddy, S. K., Swaminathan, V., & Motley, C. M. (1998). Exploring the determinants of Broadway show success. *Journal of Marketing Research*, 35(3), 370-38.
- Simonton, D. K. (2004). Film awards as indicators of cinematic creativity and achievement: A quantitative comparison of the Oscars and six alternatives. *Creativity Research Journal*, 16, 163-172.
- Stemler, S. E. (2004). A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Practical Assessment, Research & Evaluation*, 9(4), para 6.
- Stigler, G. J. & Becker, G. S. (1977). De gustibus non est disputandum. *The American Economic Review*, 67(2), 76-90.
- Towse, R. (2010). *A Textbook of Cultural Economics*. Cambridge: Cambridge University Press.



- Watson, M. R., Anand, N. (2006). Award ceremony as an arbiter of commerce and canon in the popular music industry. *Cambridge Journals*, 25, 41-56.
- Vermeulen, F., Van Dijck, M., & De Laet, V. (2013). The test of time. Art encyclopedia and the formation of the canon of seventeenth-century painters in the low countries. *Empirical Studies of the Arts*, 31, 81-105.
- Wijnberg, N. M. (2011). Awards. In R. Towse (Ed.), *A Handbook of Cultural Economics, Second Edition* (pp. 63-66). Cheltenham: Edward Elgar Publishing.

## Appendix 1

**Table 5.** *Characteristics of six book award-granting juries and their prizes*

<i>Award</i>	<i>Number of jurors</i>	<i>How are jurors chosen</i>	<i>Who can submit books?</i>	<i>Eligible books</i>	<i>Amount of works submitted</i>	<i>Prize</i>
Pulitzer Prize for Fiction	20	Board members are the jurors and makes all prize decisions for all award categories including fiction	Publishers and writers	Books written by American citizens and published in the US	Unknown	Winner: \$10,000 (= €8981.50) Finalists: unknown
National Book Award for Fiction	5	By the board members of the National Book Foundation	Publishers	Book written by American citizens	Unknown, but between 150-500	Winner: \$10,000 (= €8981) Finalists: \$1,000 (= €898)
PEN/Faulkner Award for Fiction	3	By the board members of the PEN/Faulkner Foundation	Publishers	Books written by American citizens	±350	Winner: \$15,000 (= €13472) Finalists: \$5000 (= €4491)
ECI Literatuurprijs (previously AKO Literatuurprijs)	5-6	By the board members of the Stichting Jaarlijkse Literatuurprijs voor fictie en non-fictie	Publishers	Books written by living authors writing in Dutch	Over 400	Winner: €50.000 Finalists: €500
Libris literatuurprijs	6	By the board members of the Stichting Literatuur Prijs	Publishers	Books written by authors writing in Dutch	±160	Winner: €65.000 Finalists: €2500
Gouden Boekenuil	5	Unknown	Publishers	Books written by authors writing in Dutch	±370	Winner: €25.000 Finalists: €1.000

## Appendix 2

**Table 6.** Overview of the number of books awarded/nominated by all six organizations

	American awards				Dutch/Flemish awards			
	<i>PL</i>	<i>N</i>	<i>PF</i>	<i>Total</i>	<i>A</i>	<i>L</i>	<i>G</i>	<i>Total</i>
Winning books	33	37	34	104	20	20	20	60
<i>only novels</i>	30	32	25	87	17	18	20	55
<i>male author</i>	20	24	26	70	16	19	20	55
<i>female author</i>	13	13	8	34	4	1	0	5
Shortlisted books	68	136	139	343	99	100	68	267
<i>only novels</i>	53	113	95	261	73	98	55	226
<i>male author</i>	41	79	82	202	83	74	57	214
<i>female author</i>	27	57	57	141	16	26	11	53
Total	101	173	173	447	119	120	88	327
<i>only novels</i>	83	145	120	348	90	116	75	281
<i>male author</i>	61	103	108	272	99	93	77	269
<i>female author</i>	40	70	65	175	20	27	11	58
Longlist books	-	-	-	-	151	98	87	336
<i>only novels</i>	-	-	-	-	100	92	59	251
<i>male author</i>	-	-	-	-	116	69	68	253
<i>female author</i>	-	-	-	-	35	29	19	83
Total included LLB	-	-	-	-	270	218	175	663
<i>only novels</i>	-	-	-	-	190	208	134	532
<i>male author</i>	-	-	-	-	215	162	145	522
<i>female author</i>	-	-	-	-	55	56	30	141

*Note.* PL= Pulitzer Prize for Fiction. N=National Book Award for Fiction. PF=PEN/Faulkner Award for Fiction. A=AKO-literatuurprijs. L=Libris-literatuurprijs. G=Gouden Boekenuil. PL: in 1989 two instead of three novels were shortlisted and in 2012 three authors were shortlisted (as usual) but no winner was chosen. Consequently, the total amount of winners was 33 instead of 34. N: from 1981-1983 there were separate prizes for hardcover and paperback fiction books. Both were included in the dataset, so that the total amount of winners was 37 instead of 34. A: in 2002 five instead of six nominees were chosen for the shortlist. Gouden Boekenuil: this award has no standard amount of books on their shortlist.