The Statistical Theory of Storytelling: Quantifying the Arts

Introduction

A story is an interaction of characters' actions on an arbitrary interval of time. By representing the characters as factors and their actions as levels, the audience's interpretation of the results (assumed by the null hypothesis to be caused by the factors) is representative of the state of the story at some moment in time, t (colloquially called the 'current state of the story'). Implicit in this description of a story is the use of the terminology of Dr. Genichi Taguchi, who had originally used *factors* and *levels* to describe elements of car manufacturing. Furthermore, Taguchi's use of a *Loss Function* to quantify a customer's dissatisfaction (e.g., with the quality of a product or process) will be extended to describe an audience's interpretations and subsequent expectations of a story's events.

An example that illustrates this concept is a 'plot twist', which is a colloquial expression that means a result (i.e. an interaction of the characters) that was not expected when compared to previous results. By breaking down the acts of witnessing the result, processing it, developing an expectation for the chance that subsequent events will occur, etc. we can begin to see how a plot twist can be viewed mathematically. Using the aforementioned definition of the 'interpretation of the results', each state of the story at discrete time intervals preceding the unexpected result (the 'plot twist') is consequentially defined to be an interpretation of all results up to that moment in time; therefore the summation of these interpretations over time from the beginning of the story (t_0) to the moment just before the plot twist (t_{TWIST}) gives the state of the story at the moment preceding the plot twist. However, there are some nuances in the middle that I will explain as we go along.

But for right now, I will mathematically define the state of the story preceding the plot twist as a function, S, based on another function F (the working knowledge of the story), which is itself based on another function, N (the expectation that an interpreted result will occur), which in turn is based on yet another function, g (the interpretation of an accumulation of results in time), which is once again a function of yet another function, I (the accumulation, or summation, of all results in time), which is finally a function of the results summed over time.[1] I will proceed in describing this framework from the observed results (y) and work my way back to the state of the story (S).

---

[1] All of these functions actually have running indices but I'll leave them off for now to avoid confusion. Additionally, I want to leave open the possibility that 'I' could also be $I_i = I_{i-1} + y_i$, where $I_{i-1} = \sum y_w \Delta t_w$, where the sum goes from w = 0 to w = i-1, so that all of information given by results up to the $(i-1)^{th}$ result has been integrated once the new result (the i^{th} result) has been observed. For now, I will stick to my guns and say that each new result is integrated along with all of the other ones previous to it at the same time, though this is probably not exactly how a human acquires information, especially when it is sequential. Nevertheless, I'll use my original expression for I_k (to differentiate it from the I_i I propose in this footnote) as a first-order ensatz.

Mathematical Framework

Accumulation

Beginning with the accumulation of all results up to $t = t_{TWIST}$, Equation 1 shows how the k^{th} accumulation of n results exists at the moment in time equal to the sum of all time intervals up to the n^{th} result's occurrence. The constraint on t_i in Equation 1 is tailored to the working example of a plot twist.

$$I_k = \left[\sum_{i=0}^{n} y_i \left(t_i - t_{i-1} \right) \right]_k$$
$$t_i < t_{TWIST}$$

Equation 1

Note that I_k is not the current state of the story, but rather it is an accumulation over time intervals ($t_i - t_{i-1}$) of witnessed facts (y_i's) that eventually will lead to an audience's perception, or as defined in this framework an interpretation (g_k), of the current state of the story.

Furthermore, since an attentive audience integrates[2] sequential facts of a story over time intervals that are large compared to an arbitrarily small time interval, it can be assumed that n -> infinity. Another way to look at this concept is that time is not recognizably discretized by the audience therefore any interval of time (say, 1 second or an hour) can be considered as a summation of infinitesimal dt's, provided that $\int dt > t_{LAPSE}$, where t_{LAPSE} is the time needed for a human to process a lapse of time, esp. in the case of witnessing the events of a story).[3]

State of the Story

Now, looking at the framework from the point of view of the desired function for the state of the story (S), Equation 2 shows the state of the story preceding the plot twist, indicated by t''.

$$S_k \sim \sum_{i=0}^{t''} F_k \Delta t_i$$
$$t'' = \sum_{i=1}^{n} \Delta t_i < t_{TWIST}$$

Equation 2

---

[2] (not mathematical, but conceptual, as in 'integrating a concept into one's action'; it's not my fault I speak English! Who came up with this language anyway?!)

[3] One could use this assumption of n -> infinity to claim that all sums related to indexed values can be changed to integrals, yet I keep the sums to remind us of the fact that results happen and are perceived discreetly.

In Equation 2, F_k represents the audience's working knowledge of the story generated with an interpretation, g_k, which in turn is dependent upon I_k (Equation 1). Remember that I_k is the sum of all y_i's (i.e. the accumulation of the results) over time that is fed into the 'interpretation function' g to generate the k^{th} interpretation, g_k. Furthermore, the k^{th} interpretation allows a 'working knowledge' function F to exist because a string of interpretations accumulated over time develop into a working knowledge. Thus an F_k is generated at t_i in comparison to the previous working knowledge of the story at the $(i-1)^{th}$ earlier time. But the exact relationship of S to F is not known yet because there are some processing details to consider.

Additionally, it is important to note that time is discretized such that the $(i-1)^{th}$ earlier time is separated from the i^{th} time by at least t_{LAPSE}; at this moment in time, F_{k-1} had existed due to a similar process beginning with the summation of the $(i-1)^{th}$ Δt. In other words, F_{k-1} would be created from the transformation of $\Delta t_i \rightarrow \Delta t_{i-1}$. Finally, it is important to emphasize that F_k is not equal to F_{k-1} because there is more input into the genesis of F_k than into that of F_{k-1}.[4]

Interpretation

Although its properties are somewhat dependent on the audience member, the key input to F_k is a function g_k. The defining property of the n' g_k's is that their values follow an approximately normal sample distribution, $N(\mu,\sigma^2)$, defined by $\mu(g_k)$, commonly known as the average, and $\sigma^2(g_k)$, its variance or its 'spread'.

$$\mu = \sum_{k=0}^{n'} \frac{g_k}{n'}$$

$$\sigma^2 = \sum_{k=0}^{n'} \frac{(g_k - \mu)^2}{n'}$$

Equation 3

[4] It is assumed that there exists a forward flow of time. Note that the temporal location of F_{k-1} depends on the audience's speed and ability to integrate the results into their working memory of the story. The audience's ability to integrate an input (i.e. the result giving the updated state of the story) consequentially is a function of the audience's processing speed. Although processing speed varies among audience members, the audience's attention is also vital to its ability to integrate inputs of a story; in other words, this is a quantitative way to track an audience's 'interest' in the story. Even though processing speed and attention are interacting factors within the context of neuroscience, I am considering them separately. Therefore considered independently from attention, F_k depends on the processing speed of the audience (i.e. how much of the story the person understands) and has a unique value for a given person because F_k is a function of many other neurological factors such as previous experiences and the ability to integrate that knowledge into a working knowledge of those experiences, etc.; but for now we will assume F_k to be the consequence of all previously observed results starting from $t_0 = 0$ of the story.

In the case of the story, the statistical sample of y_i's serve as the inputs once they are summed over the $(t_i - t_{i-1})$ time intervals to create I_k, which is then fed into g_k. But since I_k changes every time the audience witnesses a new result (thus the k index), there is for good reason an index on I as I've been doing.

Expectation

Furthermore, the nature of g_k is such that it interprets I_k and establishes an expectation[5] via the shape of its distribution N (i.e. the evolving interpretation of results)[6]. In other words, the sample distribution of the n' g_k's can be turned into a probability distribution function, $N(\mu,\sigma^2)$, whose frequency values predict the probability that a particular value of g_{k+1} will be generated. The genesis of g_{k+1} of course begins with the $(i+1)^{th}$ result (y_{i+1}) via the relationships I've established (i.e., Equation 1 followed by the interpretation of each I_k as its generated). Therefore as an example, since N is a normal distribution, its mean (μ) represents the most probable value of g_{k+1}, which I define as T.[7]

Now, since N is a probability function depending on each g_k, there is actually a new normal distribution (i.e. a new distribution of results expected to occur with some probability) for every observation and processing of a new result. Therefore $N(\mu,\sigma^2) \rightarrow N_m(\mu_m,\sigma_m^2)$,[8] assuming that the $(i+1)^{th}$ interpreted result, g_{k+1}, happens after g_k.[9] In other words, since the expectations of the

---

[5] Note that this is not the mathematical use of expectation but rather the noun for expecting an event to occur.

[6] Since $N(\mu,\sigma^2)$ is a condensation of the sum of the n interpretations via μ and σ^2, thereby establishing a probability for an interpreted result to occur, the working knowledge (F) of the audience is derived with N.

[7] It might seem redundant to reassign the value of μ to T, but I use T here to relate it to the idea of Taguchi's loss function, L, whereby any deviation from the μ of the $(k+1)^{th}$ interpreted result (g_{k+1}) generates an increase in L because $L(\mu',(\sigma')^2)$ is defined to be $(cost)*((\sigma')^2 + (\mu'-T)^2)$; any deviation of the mean increases L as well as any spread of the mean. Therefore each iteration of g to produce a g_{k+1}, g_{k+2}, etc. increases L if these values of g are not equal to μ' (I have primed μ and σ to differentiate them from those of the paper). However if they are equal to μ', then L decreases because $(\sigma')^2$ decreases since the spread shrinks for every g added to the histogram that falls on μ'. Thus each $(k+1)^{th}$, $(k+2)^{th}$, etc.th iteration of g changes L. Furthermore, what is sought via the interpretation of results by the audience as they are witnessed is that $\mu' \rightarrow T$; by the very nature of interpreting the results, the audience is expecting *not* to be surprised. Consequentially this allows for the audience's taste to be considered (i.e. do they like so-called 'happy' endings, do they like 'bad' endings, or are any other events of the story disturbing for them?). Therefore the model of N I am using in the body of the paper is the Taguchi Loss Function where 'cost', the proportionality constant, is equal to surprise in the context of witnessing a story given some tolerance (Δ) which I'll get to later.

[8] Note that I have made the N index m so that it is not to be confused either with the i or k indices, one of which (i) has been summed over before the computation of N and the other of which (k) is used to compute N.

[9] Once again, the reigning assumption is that time is going forward from the i^{th} to the $(i+1)^{th}$ moment.

audience are condensed into the probability distribution, and since the distribution evolves with each time interval as well, there is also a running index on N.

Returning to the concept of T, T is the interpretation of the most expected result to follow y_i[10]; in other words, it defines the peak of the g_k distribution. Thus, the distribution represents a probability as I've said of the most expected result in the eyes of the audience given the k^{th} processing of the i^{th} result. Since T is the most expected result to follow the k^{th} interpretation (g_k) of the i^h result (y_i), T is defined as the mean, μ, of $N_m(\mu_m,\sigma_m^2)$ because the mean occurs with the highest probability in normal distributions.

General Application: Clarification of N

Now let's use this rough framework on a story example: the plot twist. Using the aforementioned ideas, the plot twist can make the transition from colloquial expression to rigorously explained concept. In the process of detailing the plot twist though, the importance of N will be clarified both as a concept and as a step in the information processing.

Discussion on Expectation

In the time interval immediately prior to the plot twist, y_i has caused through the g_k's interpretation an expectation that T will remain to be equal to μ (i.e. that y_{i+1} will generate a g_{k+1}=T).[11] However, due to the plot twist (and its inherent definition as a result that is not predictable from interpretation of the results previous to its occurrence) the $(i+1)^{th}$ result that causes an interpreted plot twist can mathematically be expressed as the following[12]:

$$y_{i+1} \equiv y_{TWIST}$$
$$g[I_{k+1}(y_{i+1})] = g_{k+1} = g_{TWIST}$$
$$|T - g_{TWIST}| > 0$$

Equation 4

---

[10] See footnote #9 for a detailed explanation using the Taguchi Loss Function.

[11] Therefore the exact analog to Taguchi is that g_{k+1} is the $(k+1)^{th}$ product or process that must hit some target, T, where in the case of story-telling, T is determined by an interpretation of all results previously witnessed.

[12] Thus increasing the Taguchi Loss Function, reworked here to be the Audience Expectation Function which is proportional to a constant defined to be (surprise)/(tolerance for surprise)^2; just in case the tolerance is less than zero.

As the equations of Equation 4 show, by defining y_{i+1} as the unexpected result, g_{k+1} emerges as the interpretation of such a result. Since g_{TWIST} is not equal to T, a result that had been processed and interpreted was not expected to happen.[13]

Furthermore, the level of surprise generated by this plot twist can be found given a tolerance (Δ) around T. Now, Δ is defined such that 2Δ encompasses nearly all of N to an arbitrary degree of accuracy. Using $|T-g_{TWIST}|$, the mathematical translation of 'no %$*&-ing way!' can be stated for a tolerance (Δ) as $|T-g_{TWIST}| >> \Delta$. Furthermore, as $|T-g_{TWIST}|$ -> infinity, the more the result that caused its associating g_{TWIST} was a plot twist. If $|T-g_{TWIST}|$ had only been slightly greater than Δ, it would have been a minor plot twist, but since it is much greater than Δ it is a major plot twist (see Figure 1 below).

The added benefit of Δ is that the interest of the audience can be considered as well as its level of surprise. Let's define Δ to be 6σ, though in actuality it could be as many σ's of the distribution as desired.[14] The consequence of choosing $\Delta = 6\sigma$ is that there is a 2 out of 10^9 probability that

---

[13] Note that the English language is designed to signal a past time before another past time (represented by the preterit) by the preterit perfect. The connotation in English though is that the preterit imperfect happens for a long time before the preterit instead of how it is meant to be used as a way to describe a scenario in place when an action being recalled in the preterit occurred. The reason I am picky about this here is that both the moment the result happens and the moment the subsequent action to process & interpret that result occurs the latter moment is in a causal relationship to the former. Therefore I assume these two events to be separated by an interval of time. Furthermore, the moment the result is processed there is yet another crucial interval of time: the time during which the audience has the expectation for a subsequent result. Consequentially this time interval is t_{LAPSE}, or the amount of time it takes for the audience to witness, process and develop an expectation. I have defined interpret in the main body of the paper to mean the 'process and develop an expectation routine'. The separation of these three moments in time is crucial to the idea of quantitatively describing storytelling. However, unfortunately English does not have a verb tense that fits between present perfect and the preterit, therefore I use the preterit obstinately to say that the processed result was expected to happen in the time leading up to the next result which occurs in the present. I only address all of this since the time intervals familiar to the English language are much bigger than the ones to which I am referring.

[14] Note that I could have made my life a lot easier here by saying $\Delta = q\sigma$, where q -> infinity, thus eliminating the possibility g_{TWIST} is a 100% twist. But I'm going to leave it as an arbitrarily defined parameter so that noise conditions have some presence in this theory -- noise conditions in the sense that the audience may have missed an opportunity to observe a result (i.e. such as the famous line after a bathroom break in the theater: "Okay, what did I miss?"). I can't resist giving the alternative 'noise' condition that the person behind you or next to you in the theater could also be talking loudly, thus creating a doubly noisy condition for your viewing pleasure! Additionally this can account for whether the audience believes the witnessed event or not, thus allowing room for the common saying 'I don't %&*$-ing believe that!' or 'that's bull-&*$%'; or conversely this can account for an audience too easily duped.

the audience will not interpret the result.[15] In other words, for whatever reason (e.g., boredom or distraction, etc.), an audience member does not witness an event with a 2 out of 10^9. With Δ defined, there is a way to determine the probability that an event will be missed by an audience member; this information comes from the magnitude of $|T-g_{TWIST}|$. Therefore if $0 < |T-g_{TWIST}| < \Delta$ then the closer $|T-g_{TWIST}|$ is to zero implies that a probability still exists that g_{TWIST} will not be processed as a plot twist. Conversely, the farther $|T-g_{TWIST}|$ is from zero and closer to Δ, the more the probability vanishes that g_{TWIST} will not be incorporated into the $(m+1)^{th}$ iteration of N as a plot twist.

Recapping this argument assuming T=0: Starting from the center of the normal distribution, N, there exists a probability that the audience will miss the startling result for the next value of g. But as the value of g moves towards Δ, the chance that such a startling result is missed approaches zero. Therefore beyond the value of g=Δ, it is impossible to miss a startling result because the level of surprise increases for higher values of g.

The analog to the audience's recognition and level of surprise is the classic example of the tolerance in diameter of a manufactured bolt. If the level of surprise becomes the customer dissatisfaction and the error in the bolt's diameter becomes the value of each g in N then Δ is the tolerance for error. For diameters beyond Δ there is loss, whereas within Δ there is an increasing probability as the value of the diameter goes from 0 to Δ that loss will be incurred (see Equation 5 for more details).

---

[15] Therefore the definition of Δ in terms of the spread of N (i.e. the audience's working knowledge) is a way to account for the alertness and processing abilities of the audience, which are based on among other things its interest in the story. Thus Δ can be defined based on the genre preference, or actor preference, etc.

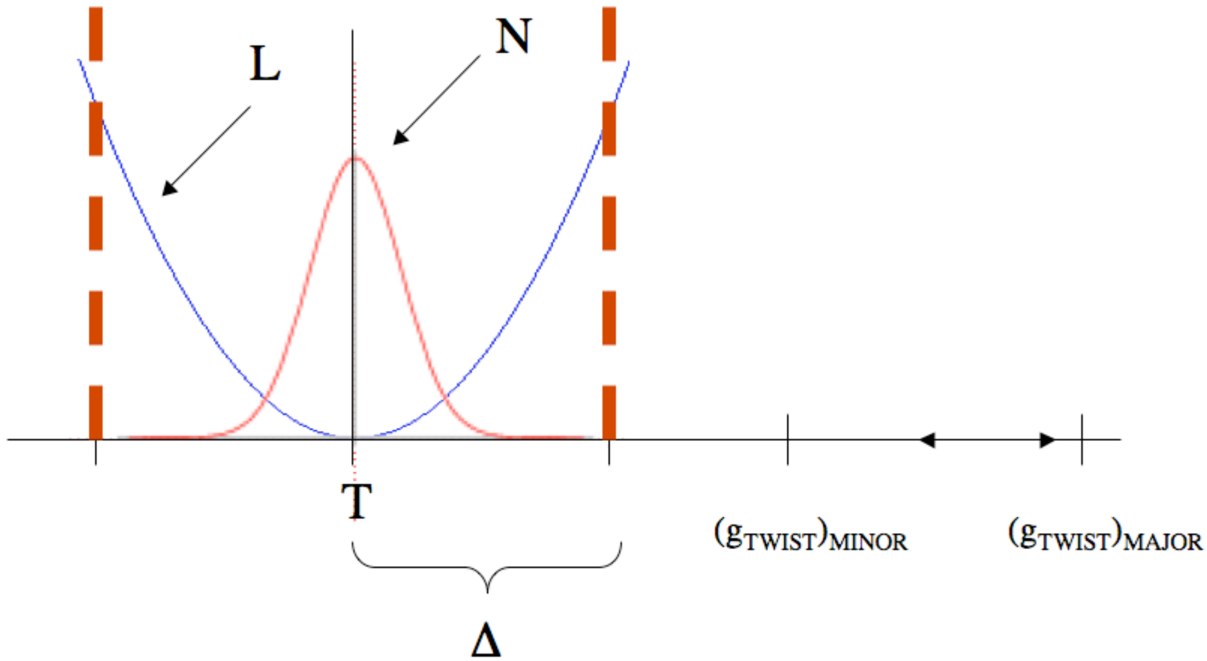

Figure 1

Above: the solid red curve shows N(g,μ,σ^2) with the tolerance limits of the expected g value denoted by the red-dashed lines at a distance Δ from T to either side. Also, the two-dimensional projection of the three dimensional Audience Surprise Function (ASF) is shown with the blue-lined curve.

Figure 1 visualizes the importance of Δ and its role in defining the audience's level, and chance, of being surprised. The Audience Surprise Function (ASF) is defined as the level of surprise and in reality its shape is a 2D parabola rotated around the vertical axis, which is defined by T. **Equation 5** defines the ASF. Note that the analogy from the manufactured bolt holds for **Equation 5** using the substitutions mentioned above.

$$ASF = \frac{surprise}{\Delta^2}\left(\sigma_{g_k}{}^2 + \left(\mu_{g_k} - T\right)^2\right)$$

$$N_m\left(g_k, \mu_{g_k}, \sigma_{g_k}\right) = \left[\frac{1}{\sqrt{2\pi\sigma_{g_k}^2}}\right]\exp\left(\frac{-\left(g_k - \mu_{g_k}\right)^2}{2\sigma_{g_k}^2}\right)$$

Equation 5

Figure 1 also shows the major and minor plot twists and it can be seen that $|T-g_{TWIST}| > 0$ for the minor one and $|T-g_{TWIST}| >> 0$ for the major one.

Mathematical Details: Closing the Gap

Now that N_m has been established as the probability of predicting the next g_k, we must tie the N_m's to the F_k's and ultimately to the S_k's. Even though I said that the F_k's are a function of the g_k's, this is actually only true once the k-index is summed over. In actuality, the F_j's (note the new index) are a function of not only the sum over the k's but they are also a function of the sum over the m's.

Working Knowledge

Each $N_m(\mu_m, \sigma_m^2)$ is a probability function that creates an expectation for the value of g_{k+1}. By defining the F_j's as a sum over the N_m's, each F_j is an accumulation over every m^{th} point of time interval during which a particular result is expected to occur; Equation 6 summarizes this point.

$$F_j = \left(\sum_{m=0}^{n''} N_m \right)_j$$

Equation 6

Therefore each F_j is a discretely occurring working knowledge of the story because it is a condensation of the observed results (i.e. F is a data processor). The processing begins upstream with the observation of y, followed by the accumulation of I, before which is the interpretation of g, whose probability of existence is approximated by N; each time this processing occurs, the working knowledge of the story becomes more refined.[16]

I use the phrase *working knowledge* because it describes how past interpretations of past results can influence pending interpretations of present results. The past interpretations of past results are the *knowledge* gained through experience and they are put to *work* to interpret present results. The idea of working knowledge is a quantitative way to account for the acquisition of knowledge through experience, also known as *learning*. In fact, I found out after writing the first draft of this paper that Antonio Damasio uses the phrase *working memory* for the same idea as *working knowledge*.[i]

In a similar fashion as I_k, the state of the story (Equation 7) can now be defined via an evolving accumulation of each F_j. In other words, the evolving working knowledge of the story accumulated over time is defined at t''. [17]

---

[16] Huge! I mean mega-implications for robotics! Come back to this one, Zephryn! Note that this is a purely rational processing of input and that both emotions and psychological factors can be put in there two by assigning weighting factors to the various stages of developing the working knowledge. For example, the interpretation, g_k, of some result, y_i, could be flawed by a 'mental block' (lack of attention?) to process the result. Or another example could be that the condensation (N_m) of the interpretation of the results could be flawed via an emotional block to converse with other people who could help with such a task because of fear of looking stupid in front of others, or whatever. Furthermore, this general process I have developed through statistics of tracing the development of a working knowledge can be used to show that integrative thinking on ALL levels (i.e. between apparently grossly different results) is far more powerful than within artificially imposed sub-disciplines.

[17] Furthermore I want to make a point about the case that I have repeated again and again that n -> infinity. One might ask 'if n goes to infinity, then how big is t_{TWIST}?' Actually, since n goes to infinity, while Δt is a finite interval fixed by t_{LAPSE}, the processing time of human audience

$$S_q = \left(\sum_{j=0}^{t'} F_j \Delta t_j \right)_q$$

$$t' = \sum_{i=0}^{n} \left(t_{LAPSE} \right)_i \equiv t_{TWIST}$$

Equation 7

In Equation 7, the gap has been closed between how S relates to F. Recalling Equation 2, it was suggested that a relationship existed between these quantities, but the definitions of g, N and F have made it made possible to relate F to S exactly.

Also, note that F_j does not exist at t'', at which time S exists, but at t'. In other words, it is not guaranteed that the information is always processed from function to function in the creation of S. Also, note that the index on Δt is *j* and not *i* because, due to the possibility of processing interruptions, Δt_i will not always be equal to Δt_j in general, though it will most likely be very close.

Understanding

In a similar way in which F is created, an understanding of the story (U) can be defined by summing over the states of the story. In other words, S contains the information of the evolving interpretations and expectations of a story's events, therefore U (Equation 8) is the audience's understanding of the witnessed results.[18]

$$U = \sum_{q=0}^{n''} S_q$$

Equation 8

Therefore, according to Equation 8, a story is only as good as its understanding.[19] In other words, a story's artistic merit is only as good as its ability to convey information to an audience, despite

---

member, the $\sum_i$ does go to infinity. But since I have studied physics for 4 years at the University of Chicago I am going to define t_{TWIST} to be at a greater value of infinity than the one produced by the $\sum_i$. Furthermore, each subsequent point in the story (i.e. every time a result occurs) occurs at some value of infinity greater than the previous in the dimension of time. I guess it could super-rigorously be done with n -> N, where N is large, but that is just as artificial as me having various 'degrees of infinity'.

[18] Note that iteratively applying this process to robotics involves using the stories as the results, as described in the context of the story; mathematically this means that $S_j \rightarrow y_i$. Thus over time, the robot gains empirical knowledge with each iteration in which S -> y.

[19] Note that U could be given an index as well thus accounting for the understanding of one audience member for every different value of the index. Consequentially, an understanding for

the audience's level of attention. A familiar example of this fact is the use of movie trailers to advertise an upcoming film. By exhibiting material that has the possibility of spurring an attentive audience (because the trailers are so short in terms of time), a potential filmgoer generates a preliminary understanding (U) of a story each S_q generated by witnessing the events in the trailer. However, due to the outlined nature of the actual function S_q, the filmgoer's interest may or may not lead him/her to see the movie.

Specific Application

Applying this theory to one example that helped in its creation, I wish to address the Damien Rice song, "The Blower's Daughter", which is featured in the movie *Closer*, which is equally well-suited for a statistical story-telling analysis. The storytelling of this song follows what seems like a love affair from afar. The refrain, "I can't take my eyes off of you" evolves as the song progresses (note the elapse of time between results, i.e. the refrain, happening) into the refrain "I can't take my mind off of you' towards the end of the song. Furthermore, for nearly 98% of the running time of the recording, the song is a story of an impassioned love affair from afar without any fulfillment. However, in the last seconds of the song the following lyrics are nearly inaudibly sung: "till I find somebody new." The addition of these last five words, which occur over the span of a few seconds, retroactively changes the meaning of the whole content of

---

the story (U) can license that person to pass judgment on it (whether they understand the story on its multiple levels or not) such that the mapping of U to another function V is possible, where V takes U as its input and determines a like or dislike of the audience member. Once again, attaching an index to V allows the same set of normal sample distribution statistics that we took of the g_k's to produce the N_m's. Therefore a V_s can be used in the expressions for mean and variance (I'll call them η instead of μ and ρ instead of σ, respectively) to generate a distribution (Z) of the V_k's, where according to the definition of V_s, η is the V_s such that it is the most probable to exist. Therefore more people liked the story (S) that generated the understanding (U), which in turn generated either the approval or disapproval (V) of the understanding of that story. Therefore the popularity of the story is only as great as its understanding (U), which must follow through all the steps I have laid down in this statistical theory of story-telling. More colloquially, if no one understands it no one will like it. A corollary of this deduction is that forms of story-telling that are 'popular' are not necessarily appealing to everyone but instead are appealing to a good number of everyone, where that 'good number' occurs at the mean (η) of the distribution of the V_s's. An example of this phenomenon is the romantic comedy or the action/adventure movie, which is not necessarily a good piece of story-telling (in fact, it is usually not, but that's a biased comment and thus not a V_s as likely to occur when compared to the sample (i.e. the worldwide audience exposed to such movies) but rather a most easily understandable story-telling. Furthermore, there are many inputs to the determination of what is most popular (one could consider a sum over the reactions to one movie by the population and compare its distribution to that of others, i.e. compare the $Z(\eta,\rho^2)$'s of all stories), including societal influences (some of which from the point of view of culture could be generated by the distribution of Z's for all stories) that are not necessarily (and most often not) separate from the cultural values of works of art, such as a story, or the society as a whole, neither of which are wholly separate (and rarely are) from one another.

the nearly five minute song. Therefore the whole purpose of the song lies within those few seconds that could either not be noticed, not be heard, or not be understood within the framework of the song. For whatever reason, any one of the steps I have laid out -- the observation of the result (y), the integration of that result (I) in terms of all previous results, the interpretation (g) of that result such that an expectation (N) develops a working knowledge (F) of the states of the story (S) ultimately leading to an understanding (U) – could be faulty and the whole point of the song on that level of story-telling could be missed.

Extensions of the Theory

Furthermore, the statistical theory of storytelling applies to any genre of art that exists temporally as a succession of results caused by factors at particular levels. For example, in music the audience exists in the same role as it did for story-telling, the results are the actions of the instruments or voices controlled by the musicians (equal role as the actors), who are the factors, operating at certain levels (i.e. performing certain pitches at certain timbres, dynamics, rhythms, etc.), whereby the levels of the factors interact to produce a result (i.e. a musical moment in time) that when summed over many musical moments in time becomes a musical event.[20] Similarly an accumulation of musical events causes an evolving chain of musical preference as before with the storytelling. Applying the statistical theory of storytelling to music allows me to address a common phenomenon known as 'rediscovering music'. Historical examples exist from Chuck Berry adopting the blues into 20th century American rock 'n roll to Felix Mendelssohn popularizing the long-dormant music of J.S. Bach in the public music scene of 19th century Europe; these are not unrelated events at all, but rather a predictable scenario that has happened many times before and will continue to happen in the future, as long as the exchange and criticism of such temporally[21] existing art can do so freely in the future. Furthermore, a corollary of this is that a particular piece of music[22] will self-propagate as a function of its availability and multiple levels of understanding. If it's simple to understand once and becomes boring after that then the music will die a quick death in the eyes of the audience, but if it is understandable once on one level (perhaps a more common level, so as to get the proverbial 'foot in the door' of the audience) and then again on another, and again on another, etc., the nature of free criticism and exchange of music can allow the rapid growth of the popularity of the piece of music once integrated over time. For this reason as a demonstrating example, Mozart has more fans now than he did when he was alive[23] because not only is his music understandable on many levels, but it also has become more prevalent as people are allowed to interact with it in its various forms (i.e. listening, playing, reading, etc.).[24]

---

[20] Once again, the different definition of I_k here can account for different ways of processing information.
[21] I say temporally existing to eliminate visual art because that cultural taste evolves over longer time scales and involves different scenarios of observing results. Instead I say 'temporal' to include the performing arts, such as music, dance, theater, etc.
[22] And I imagine this is applicable to all package-able performing arts media.
[23] I'd say he actually has more fans now in just the U.S. alone than he had when he was alive in the solitary environs of Europe.
[24] Thus the common saying, "How do you get famous? Die."

Conclusion

While there may be many extrapolations of this exciting theory, which for the first time maps art into scientific thought in a way to account for trends of culture (also known as style), I will close this paper with thanks not only to the genius of Damien Rice (decidedly so because of his now quantifiable way to alter the understanding of a story within a few seconds, thereby changing the understanding of 4 minutes and 48 seconds of story-telling without saying a word during that time interval but rather *after* it) but also the geniuses who have created every piece of work that intuitively led me to follow my nose on such a powerful argument as this theory.

---

[i] Damasio, Antonio. *Looking for Spinoza: Joy, Sorrow and the Feeling Brain*. Harcourt Inc.: New York, 2003.