Can a car really run on empty?: An experiment testing for within- and between-individual comparison classes in relative and absolute adjectives.

Haram Kang

6004466

August, 2021

Table of Contents

Abstract

This thesis investigates two theories that provide an account for the distinct difference between relative and absolute gradable adjectives. Kennedy (2009) on one hand proposes that the difference is fundamental to the underlying scale structure of the adjective, while Toledo & Sassoon (2011) on the other hand proposes that the difference is fundamental to the type of comparison class that the adjective is assigned through their typical arguments. This thesis reports on an experiment investigating the nature of the type of comparison class for relative and absolute adjectives, as well as their dependency on their scale structure for interpretation. I find that a combination of the two theories must be made in order to fully comprehend the result of the experiment, and further propose that functionality is one pragmatic factor that influences the interpretation of positive forms of gradable adjectives, among other possible factors.

1. Introduction

The topic of this thesis is the semantics of the unmarked positive form of gradable adjectives. The goal of this thesis is to provide experimental findings using a new experiment to test if the semantics of all gradable adjectives is context dependent, and if so, explore which context factors are relevant in interpreting the semantics of positive forms of gradable adjectives.

Gradable adjectives are adjectives in which the properties that they evaluate have the quality of being gradable. Take (1) for example:

(1) *Usain Bolt is fast*.

Gradable adjectives can be defined as "adjectives that support orderings of the objects in their domains relative to some scalar concept" (Leffel, Xiang, & Kennedy, 2016). *Fast* as seen in (1) is an example of a gradable adjective, as the property that is relevant to evaluate this statement, speed, is a scalar concept: it is possible to take two entities, Usain Bolt and Tyson Gay, for example, and order them on a scale according to their speed to determine if one has higher speed than the other, and therefore faster than the other. Some examples of non-gradable adjectives are adjectives like metallic or excellent, which are not associated with a scalar concept. Gradability of gradable adjectives such as *fast* can be easily identified by the fact that the adjective can take the comparative and superlative forms such as *faster* and *fastest*. The existence of these forms tells us that the adjective is a gradable one, because the interpretation of comparative forms by nature utilizes a scalar concept: when something is

faster than something else, it means that it displays a higher degree of speed than the other on a conceptual scale of speed.[1] Since gradable adjectives utilize a scalar concept, when the property is gradable, the intensity of the property can be represented as a point (or a range) on a scale, such as represented in Figure 1:

Lower speed   Higher speed

Figure 1. Visual representation of the scale of speed. The scale is used in order to determine whether a subject has the property of being *fast*. The higher the subject is ordered on this scale, the faster it is compared to entities that are ordered lower on the scale.

It is possible to refer to speed as a specific point on this given scale. This is referred to as a degree of speed, and it is the collection of such degrees that constitutes a scale.[2] The degree of speed is directly related to the property of being fast, as the definition of being fast is displaying a high degree of speed. Therefore, gradable adjectives which represent a scalar concept such as *fast* and *slow* can be understood as degree relations between degrees and individuals.[3] This general relation between degrees and individuals is the meaning of the adjective itself, but the meaning of its positive form involves a more detailed relation that utilizes specific degrees (definition of positive form is explained in section 2). For example, *fast* in (1) is a positive form of the adjective and it denotes the relation between Usain Bolt and specific degrees of speed: (1) is true if speed displayed by Usain Bolt includes degree *d*, given that *d* is a minimum degree of speed that qualifies an individual to be considered fast. In other words, the maximum speed of Usain Bolt is a degree somewhere on the scale given in Figure 1, and (1) is true if that degree is at least the degree *d*, the degree of speed one needs to display to be considered 'fast' – whatever that degree might be. Its comparative forms such as *faster* can also be understood as denoting similar degree relations. *Usain Bolt is faster than Tyson Gay* is true if the maximum degree of speed of Usain Bolt is higher than the maximum

---

[1] Although all adjectives with comparative forms are gradable, not all gradable adjectives have comparative forms. *Beautiful* is an example.

[2] It is easy to visualize the scale of properties like speed as a collection of degrees, as boundaries of each degree can be defined mathematically. However, this is also not a requirement of a gradable scale. Adjectives such as *wise*, for example, are multi-dimensional in the sense that there is no single element which constitutes the degree of the adjective. Although there is no clear-cut, distinct guidelines to separate one degree from another, one can be wiser than another, meaning that there can still be an arbitrary standard that distinguishes one degree of wisdom to another. Therefore, a scale such as one shown in Figure 1 would still be applicable to such adjectives.

[3] The formal type of gradable adjectives under this view would be <d,<e,t>>. There are proposals for a different type for gradable adjectives, but that is outside the scope of this thesis.

degree of speed of Tyson Gay on the scale of speed. In this way, the semantics of both positive forms and comparative forms of gradable adjectives utilizes a linearly ordered scale composed of a set of degrees.

The structure of an adjective's scale can differ depending on the adjective, however. This difference in the structure of adjectives is something that can be recognized intuitively, as it will be discussed more in depth in Section 2. For example, compare the adjective *full* to the adjective *fast*. Although they are both gradable adjectives (one can be fuller than another), intuitively we can feel that how we understand *full* is different from how we understand *fast*: something being *full* seems to provide more concrete information about its subject rather than *fast*. When something is full, we know that it means that the subject is filled to the maximum; but when something is fast, it does not mean that the subject is speedy to the maximum – out of context, we do not know what something being fast really entails. Kennedy (2007) provides compelling empirical evidence to reasonably believe that there are two different types of gradable adjectives which differ in their semantic behavior, so-called 'absolute' adjectives such as *full* and so-called 'relative' adjectives such as *fast*, and that this fundamental difference between the two types arise from them having an innately different scale structure. Kennedy proposes that although they seem to perform fundamentally the same semantic function, and that there are clear reasons to believe that both are gradable adjectives, relative and absolute adjectives inherently differ in their scale structure and therefore we interpret them differently. Further explanation of Kennedy's claims is presented in Section 2. In short, the difference between relative and absolute adjectives according to Kennedy is that the truth conditions of the positive form of relative adjectives are context-dependent, while it is not context-dependent for absolute adjectives but rather solely dependent on the lexical semantics of the adjective. He proposes a general principle for gradable adjectives that he calls *Interpretive Economy*, which ultimately explains why relative adjectives are by default context-dependent while absolute adjectives are not.

Although the principle provides a great fundamental idea that interpretation of absolute adjectives by default rely on the lexical semantics of their adjectives rather than context in deciding the standard of comparison, it does not provide specific conditions of how this default preference gives way to context-dependency, as empirical evidence show that it sometimes does. A great deal of the thesis was inspired by Toledo and Sassoon (2011), who in summary claim that unlike Kennedy's proposal, both absolute and relative adjectives are context-dependent when it comes to their interpretations. Later sections will discuss ideas from Toledo and Sassoon in detail, the impact it had on the design of the experiment, and its connection to ideas presented by Kennedy. The purpose of this thesis is to provide experimental data to put to test the validity of Kennedy and Toledo and Sassoon's proposed theories regarding positive forms of gradable adjectives and use its results to further expand their proposed theories in detail, such as expanding the definition of Kennedy's *Interpretive Economy*.

The structure of the thesis is as follows: In Section 2, more relevant concepts and topics regarding the interpretation of positive forms of gradable adjectives are introduced and

explained. In Section 3, the details of the experiment and its results are presented. Section 4 contains relevant discussion of the results of the experiment and its connection to prior theories regarding gradable adjectives. Section 5 will discuss what we can take from the results of the experiment and summarize its relevancy to the general field of semantics and pragmatics.

2. Theoretical Background

2.1 Positive forms of gradable adjectives

Kennedy (2007) points out that it is a bit paradoxical that the unmarked, positive forms of gradable adjectives which are morphologically simple, appear to be more complicated to characterize in terms of a compositional semantic analysis compared to more morphologically complicated forms of the adjectives. To understand the semantic complexity of adjectives in their positive forms, we must understand how they are used differently than their modified forms. Take (2) for example:

(2) Usain Bolt is very fast.

Very is what is called a degree adverb, a type of adverb that modifies the degree of a property. More specifically, *very* is considered an intensifier, as it intensifies (increases) the degree of the property that it modifies. We can understand *very* to establish an ordering relation between two degrees: the predicate *very fast* in (2) is true of Usain Bolt if the degree of his speed not only meets but is considerably higher than degree *d,* a minimum degree of speed that is required to be considered fast. In other words, what makes him 'very fast' is that he is much faster than 'fast'. We assumed the semantic type of gradable adjectives to be <d,<e,t>> in the previous section, and this relates well to the fact that these gradable adjectives are easily modified by degree adverbs such as *very*: gradable adjectives which are of type <d,<e,t>> combine with degree denoting morphology to produce a property of individuals. However, positive forms of adjectives do not combine with any degree morphology, such as in (1), which is repeated here:

(1) Usain Bolt is fast.

How is *fast* able to denote a property of individuals without combining with any degree modifiers such as *very*? A popular solution to this problem, which is also adopted in this thesis, is assuming that gradable adjectives carry a positive null morpheme that has the same semantic function as an overt degree morpheme (Cresswell, 1977; von Stechow, 1984; Kennedy, 2007). This null morpheme, called POS, is assumed to be the component within the semantics of the positive form of a gradable adjective that provides a standard degree of comparison for the adjective. Under this understanding of POS, the main question now is exactly how this null morpheme provides a degree standard to the adjective. For example, as exemplified by its usage in (1), the adjective *fast* denotes the property of having a degree of speed which is above degree *d,* a certain standard of speed. The question is how this standard of speed is determined.

2.2 Comparison classes and standards of comparison

Before further discussion of how a standard is determined, it is important to note that there are two concepts related to comparison that will be discussed: comparison class and standard of comparison. These two concepts are related to each other but are different concepts. The general assumption is that the standard of comparison is always computed relative to a comparison class (van Rooij, 2011). In other words, the standard of comparison is derived from a comparison class. A comparison class is a set which contains all individuals that the subject of the comparative predicate is being compared to. The standard of comparison is a specific degree on the scale of the adjective that is derived for comparison, using the set of individuals in the relevant comparison class. Take (3) for example:

(3) Usain Bolt is fast for a male human.

For a male human in (3) is known as a for-PP, and it explicitly introduces a comparison class in (3). *For a male human* sets the comparison class so that the members of the comparison class are restricted to male human beings and includes all male human beings. That means we are comparing the speed of Usain Bolt to the speed of other male human beings, which are the members of the comparison class in this given case. From this comparison class we derive a standard of comparison, which is the actual degree of speed which we compare with speed of Usain Bolt to determine if he is indeed fast for a male human. A quick Google search shows that the average running speed for a man is about 12 km/h. If being above the average speed is what we are using to determine if a certain speed is fast for a male human, for instance, then this speed of 12km/h would be the standard of comparison that we would derive from the given comparison class: if a male human runs at the speed of 12km/h or faster, then he is considered fast for a male human being. Another quick Google search shows that Usain Bolt reached the speed of 44.72km/h during his record 100m dash, so against the standard of comparison of 12km/h, Usain Bolt would definitely be considered fast, and therefore (3) would be considered true.

Now let us look at (4) in addition to (3) to discuss further how a comparison class can be explicitly stated and manipulated:

(4) Usain Bolt is a fast human, but he is not fast.

In (3), it is stated that the comparison class that Usain Bolt is being compared to is a set of other male humans, and therefore the standard of comparison is a degree of speed that we derive from this comparison set that we believe would best exemplify the comparison class (the average speed, for example). Therefore, we concluded that (3) is true because Usain Bolt is indeed fast compared to the standard derived from the selected comparison class. However, things are different with (4). Those who agree with (3) can also agree with (4), although the conclusion of the statement is that Usain Bolt is not fast. Anyone who knows how fast Usain Bolt can run knows that he is a very fast individual, but under a different comparison standard, it may be agreed upon that he is not fast. Compared to an airplane, a comet, or the speed of light, for example, Usain Bolt's speed may seem much less impressive. Thus, the reason why (4) is ultimately not contradictory is because the comparison class in this case is

not limited to male humans as it was for (3). The first usage of *fast* in (4) utilizes a comparison class consisting of all other human beings, as *fast human* implies. The implied comparison class for the second occurrence of *fast* in (4), however, is a different comparison class that consists of non-human things which include objects with immense speed. If the set of the second occurrence of *fast* includes all things that display speed such as an airplane, then the standard of comparison that is derived from this comparison class is much different than the one chosen for (3) or the one chosen for the first occurrence of *fast* in (4), and it ultimately influences the truth value of the sentence. What can be concluded from these examples is that the truth condition of an assertion involving a gradable adjective depends on its standard of comparison derived from a given comparison class, which can be explicitly stated through linguistic means such as a *for-PP*. However, the question still remains with (1): most people who read sentence (1), if they know who Usain Bolt is, will agree to the statement as he is a fast human being. However, (1) does not explicitly introduce a comparison class like (3), nor does it introduce other predicates to imply other comparison classes like (4). So how do people construct a standard of comparison for (1), so that we can use it to determine the validity of the statement?

How we obtain a standard of comparison for positive forms of gradable adjectives is typically explained very simply in the literature: it is determined by context. Juxtapose (5) and (6) along with (1), as repeated below, to understand what it means for the standard of comparison to be determined by context:

(1) Usain Bolt is fast.
(5) The 6-year-old is fast.
(6) The airplane is fast.

Although all three sentences describe the subjects of the sentence as fast, they all display different thresholds for what is considered fast because the subjects of these sentences are different, and it is these subjects that determine the threshold for what is considered fast. We use our world knowledge to determine the validity of these statements. If a 6-year-old ran at the speed of 12 km/h, which is the average speed of an adult male, many will consider (5) to be true. However, if an airplane flew at the speed of 12 km/h, no one would consider (6) to be true. Likewise, if Usain Bolt as an Olympic athlete ran at the speed of 12 km/h, people would also consider (1) to be false. Under normal circumstances, however, we do not determine whether an airplane is fast by comparing it to the speed of a running man or determine whether an Olympic athlete is fast by comparing their speed to the speed of an average non-Olympian human being. For adjectives like *fast*, we most commonly use a comparison class that consists of other individuals of the same kind as the subject: we compare the speed of an airplane to the speed of other planes, and the speed of an Olympic runner to the speed of other Olympic runners. This example shows that standard of comparison for many gradable adjectives are context-dependent, such as being dependent on the subject of the sentence. However, this is not the case for all gradable adjectives. The standard of some gradable adjectives such as *full* does not require context to be determined. The difference between these adjectives is discussed more in depth in later paragraphs. What is important in this section is that we use context and world knowledge to create a comparison class for

adjectives if it is required.

2.3 The role of POS

We now know that we use world knowledge and context to create a comparison class when the comparison class is not explicitly stated in an assertion. Now we turn to the actual mechanism of the positive form of gradable adjectives that allow these adjectives to be context dependent and allows them to derive a concrete standard of comparison from a given comparison class. The function of deriving the standard of the comparison is generally assumed to be held by the POS of the positive form. Kennedy, following Bartsch & Vennemann (1972), provides a possible denotation of POS that could explain its mechanism, which is I adopt below in (7):

(7) $[[_{\text{Deg}} \text{POS}]] = \lambda g \lambda k \in D_{<e,t>} \lambda x. \breve{g}(x) > \textbf{norm}(k)(\breve{g})$

ğ is a measure function that corresponds to the maximum degree of the degree relation g (the adjective). ğ(x) therefore returns the maximum degree of the relevant property displayed by the individual *x.* The notation **norm** is "a function that returns the average degree to which objects in the set defined by *k* (the comparison class) measure ğ". The **norm** in POS will compute the average degree of the comparison class, which can either be explicitly stated as in (3) or derived from context when it is not made explicit such as in (1). In this way, the function of POS does not change whether the comparison class is explicit or not (at least in the scope of this thesis). What is important is that the mechanism of deriving the standard of comparison exists and functions identically no matter if the comparison class is explicit or not. The question now is exactly what influences the determination of the comparison class for positive adjectives. In other words, which contextual factors decide the comparison class for gradable adjectives when the comparison class is not explicit? We saw with (5) and (6), for example, that the kind of the subject used with the adjective influences the comparison class of the adjective. Are there any other contextual factors that influence the composition of the comparison class as well? This question lies at the center of this thesis, and in Kennedy (2007) and Toledo & Sassoon (2011) as well. Their reasoning and conclusion are discussed later in this chapter and in later sections. But before doing so, we must also discuss the concept of vagueness. In the next paragraph, we first review the implementation of comparison class and standard of comparison in the interpretation of gradable adjectives, while assuming (7) to be true of POS.

For (3), *Usain Bolt is fast for a male human,* the for-PP *for a male human* establishes that the comparison class set for Usain Bolt in this context is a set of all male humans. As **norm** of POS takes the average of ğ (the maximum degrees of speed) displayed by all members of *k* (set of all male humans), the truth value of the statement depends on whether the degree of speed of Usain Bolt exceeds the average degree of speed of male humans. For (4), *Usain Bolt is a fast human, but he is not fast,* the negation statement *but he is not fast* establishes the comparison class for Usain Bolt in the second clause of this sentence to be a set not

containing all humans, but a set containing all other beings in the world that display a property of speed. POS takes a different *k* (set of all things with speed) in this case, and the truth value of the statement depends on whether the maximum degree of speed of Usain Bolt exceeds the average maximum speed of all other beings with speed in the world. For (1), *Usain Bolt is fast,* the comparison class is set implicitly through context. When we usually set out to determine if the speed of an Olympic athlete is fast, we compare their speed to other individuals of its kind, in this case other Olympic athletes. Therefore, the comparison class for (1) would be a set containing all other Olympic runners, and the truth value of (1) depends on whether the maximum degree of Usain Bolt's speed is greater than the average maximum speed of other Olympic runners.

At this point, one may find the **norm** function within POS to be questionable. If **norm** is truly part of the semantics of POS, then every time we derive a standard of comparison from a comparison class which either contains too many individuals (i.e., all male humans) or individuals which we do not know the averages of (i.e., all things with speed), it would either take us years just to compute the average of a comparison class or simply not be able to come up with an average. In the literature, researchers abandon the **norm** and replaces it with a function that does not compute an average but computes a degree which marks the point of "standing out" (Graff, 2000; Bogulawski, 1975; Kennedy, 2007). What this means is that instead of deriving a standard of comparison from a set of a comparison class by computing the degree which represents the average of all individuals of the comparison set, it computes a degree that when an individual in the set displays a degree of the property above this degree, it causes them to "stand out" among the other individuals in the set. How we cognitively recognize something as "standing out" is not relevant for this paper, but the important part about it is that by putting this function, which we call **s**, in place of **norm** in the semantics of POS, we can explain how we are able to derive a standard of comparison from a large or an unknown set of comparison class.

(8) $[[_{\text{Deg}} \text{POS}]] = \lambda g \lambda k \in D_{<e,t>} \lambda x.\breve{g}(x) > \mathbf{s}(k)(\breve{g})$

With this new function in place, as shown in (8), we compute assertions with positive forms of gradable adjectives similarly but in a slightly different manner than when we assumed POS contained **norm**. For instance, when we look at (3) where we compare Usain Bolt to all male humans, we do not compute the average maximum speed of all male humans but rather the degree of speed which would mark a male human to "stand out" (with respect to their speed) if the maximum degree of speed they displayed is above this degree. To consider if (3) is true, then, we would see if Usain Bolt's maximum speed is equal to or above this degree. This new **s** function within POS now gives us a satisfying semantics of POS that can derive a standard of comparison from both implicit and explicit comparison classes, as well as comparison classes whose set is large or unknown.

2.4 Three traits of vagueness

Now we can finally discuss the central question regarding which context factor influences the determination of implicit comparison classes for positive forms of gradable adjectives. To start answering this question, we first start with the observation that many positive forms of gradable adjectives are vague, while many others are not. There are three properties that make a predicate vague: truth conditional variability, Sorites paradox, and borderline cases. These three phenomena are very briefly discussed in this section.

Truth conditional variability is the phenomenon where the truth condition of a statement varies due to a shift in the standard of comparison. We saw that (1) is true when Usain Bolt is compared to other humans but compared to other things in the world that display speed (1) could be concluded to be false. Although (1) stayed the same, the shift in the standard of comparison due to a different comparison class alone caused the truth value of the statement to differ. This is the first phenomenon of vague predicates. The second phenomenon is called the Sorites paradox, illustrated below:

(9) *The Sorites Paradox*
    P1. Any person running at 45 km/hour is fast.
    P2. Any person that runs 0.1 km/hour slower than a fast person is fast.

    C. Any person running at 1km/hour is fast.

Propositions P1 and P2 both seem valid at first glance, but the conclusion that arises from the two propositions is obviously false. If the conclusion is false, then it must be that there is a problem with one of the propositions. However, it is difficult to pinpoint exactly what the problem is. Even if we were to identify the problem, which most likely resides in the reasoning of P2, it is also unclear why P2 looks so unproblematic in the first place. This paradox is another hallmark of a vague predicate containing a gradable adjective. The last phenomenon of vague predicates is borderline cases. If Usain Bolt runs at 45 km/h, he is clearly fast. Another person that runs at 3km/h is clearly slow. What about someone who runs at 15km/h?[4] For all vague gradable adjectives, there is a grey area of membership where it is difficult to decide whether the member can be considered to have the property of the adjective or not. Additionally, not only is there uncertainty about whether a degree has the property of the adjective, there may also be uncertainty about whether a degree is classified to be a borderline case or not: some may consider someone running at 15km/h fast, some may consider the same person slow, and others may view it as a borderline case. In this way, borderline cases bring a higher order of vagueness, and it is the third sign of vague predicates.

These three phenomena can be found for all vague adjectives, but not all gradable adjectives. Take for example how we use the gradable adjective *empty* when describing a jar full of

---

[4] Speed of running may not come natural to a person, but the same factor applies to other gradable adjectives. Kennedy uses price of coffee to demonstrate that the adjective *expensive* is vague: $1.50 per pound of coffee is clearly not expensive, and $20 per pound is clearly expensive, but $9.25 per pound is unclear and is considered to be a borderline case.

marbles. Once we start taking the marbles out of the jar one by one, it is very clear when the jar is empty: when the last marble is removed from the jar. There is no truth conditional variability, sorites paradox, nor borderline cases when it comes to such adjectives. This leads us to conclude that not all gradable adjectives are vague, but only some.[5] This division of vague and non-vague adjectives is the driving factor for the division of two classes of gradable adjectives for Kennedy, and he uses this distinction between gradable adjectives to answer his central question.

2.5 Two types of adjectives: Kennedy (2007)

Kennedy classified adjectives which do not display vagueness, such as *empty,* as absolute adjectives: these are adjectives that do not display truth-conditional variability, Sorites paradox or borderline cases. Let us take another adjective used in sentence (10), *dirty,* as an example:

    (10)   The table is dirty.

We can see that the adjective *dirty* does not show the three signs of vagueness. A dirty table is dirty if it has any amount of dirt on its surface – the truth condition of its definition does not vary depending on a contextual standard of comparison. It does not display any case of the Sorites paradox, as there is a clear degree in which a clean table becomes a dirty table (any minimum degree of dirt). The clear degree between a clean table and a dirty table also means that there are no borderline cases. Therefore, Kennedy classified *dirty* as an absolute adjective. This is not the case with adjectives that do display vagueness, such as *fast* as I showed in section 2.4. Kennedy classified adjectives that display vagueness, such as *fast,* as relative adjectives.

Although the distinction of relative and absolute adjectives existed before Kennedy, he found that this distinction was not addressed properly in the literature in explaining to which extent our interpretation of gradable adjectives was context dependent. A central connection that he makes is the connection between the distinction of relative and absolute adjectives with their scale structures. (11) visually represents the four possible scale structures that a gradable adjective can have:

    (11)   Open Scale

                Partially Closed Scales

                Totally Closed Scale

---
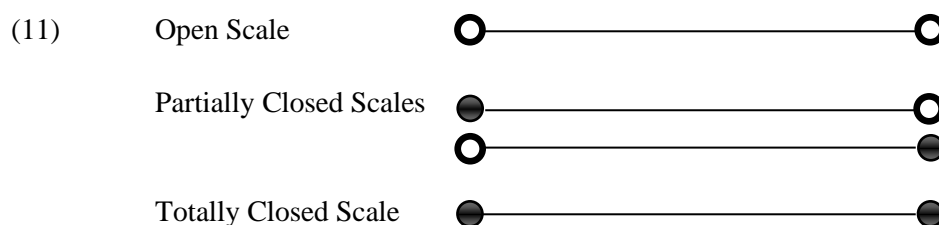
[5] It is also important to note that vague predicates are not always vague. For instance, a sports team can say that you must be fast to try-out for a spot on the team, and the cut-off point is 15km/h. A vague adjective *fast,* in this case is made non-vague. The difference between vague gradable adjectives and non-vague gradable adjectives is that non-vague gradable adjectives do not have the option of being vague, while vague gradable adjectives have the option of being not vague.

A scale is considered an open scale if the underlying scale structure has no end points on either end: this makes the scale open. How this scale structure translates to the behavior of the adjective is that for adjectives with this open type of underlying scale structure, the only requirement for an argument to potentially display the property of the adjective is to display any degree of the relevant concept. Take for example again the gradable adjective *fast,* which has an open scale. As long as an object displays any quality of speed, there exists a context in which the object can be considered fast: as we saw, the standard of comparison for *fast* can change depending on context, which can change the comparison class used with the adjective. It is possible for any object to be considered fast when displaying any degree of speed: anything displaying any degree of speed could be considered fast compared to something that does not display the quality of speed at all, such as a snail moving at the speed of 0.1km/h compared to a stationary rock. This is the case with all open scale adjectives, as their standards seem to be purely contextual.

This is in contrast to adjectives with a closed scale, where the scale can be a partially closed scale where it is either closed at the bottom of the scale (low endpoint) or at the top of the scale (high endpoint), or it could be totally closed and be closed at both ends. The requirement for the argument to potentially display the property of closed scale adjectives is to either display the highest degree of the relevant property or zero degree of the relevant property. Take for example the adjective *dry*, which is considered to have a closed scale with a high endpoint. This means that the requirement for something to be dry is to display a maximum degree of dryness: whatever the object is, for it to be considered dry, it must not display any degree of wetness, hence the maximum degree of dryness. [6] This is unlike adjectives with an open scale structure, whose standard can change depending on context.

The widely accepted empirical method to determine which scale structure underlies which adjective is to test which type of degree modifiers can be used with the adjectives (Rotstein & Winter, 2004). There are many different types of different degree modifiers, depending on the semantic role they have. Rotstein & Winter predicted that seeing which adjective can be paired with which type of degree modifier should reveal if high or low endpoints are present in the underlying scale structure of the adjective. The two relevant classes of degree adverbs here are degree adverbs known as 'maximizers' and 'diminishers', terms taken from Paradis (1997). Maximizers are degree adverbs such as *absolutely*, and they seem to indicate a sense of completeness. Diminishers are adverbs such as *slightly,* and they are used to indicate a low point on a scale. Using these adverbs, it was tested whether validity of pairing with these adverbs can predict which adjectives have a low endpoint (also referred to as a minimum standard adjective, as the adjective utilizes its low endpoint as its standard of comparison), a high standard (also referred to as a maximum standard adjective), neither or both in their

---

[6] The adjectives *dry* and *wet* share the same scale – *wet* is a minimum standard adjective because it only takes a minimum degree of wetness in order for something to be wet. *Dry* is a maximum standard adjective because it takes the maximum degree or dryness for something to be dry – this degree where something is considered dry is on the same scale used for wetness and can also be seen as zero-degree of wetness.

scale structure. An example of the test is shown in (12) – (15):

(12)   Open scale adjectives: ??absolutely fast/??slightly fast

(13)   Minimum standard adjectives: ??absolutely wet/slightly wet

(14)   Maximum standard adjectives: absolutely dry/??slightly dry

(15)   Totally closed scale adjectives: absolutely opaque/slightly opaque

(12) – (15) show that the adverbs are indeed useful in showing whether an adjective is open or closed. It shows that open scale adjectives cannot be paired with neither maximizers nor diminishers, while totally closed scale adjectives can be paired with either. It also shows that minimum standard adjectives cannot be paired with maximizers but can be with diminishers, while maximum standard adjectives cannot be paired with diminishers but can be with maximizers.

The central claim that Kennedy makes in his proposal is that the distinction between absolute and relative adjectives, in other words why some gradable adjectives display vagueness, and some do not, fundamentally depends on which kind of scale structure underlies a given adjective: absolute adjectives have a closed (either partial or total) underlying scale, while relative adjectives have an open underlying scale. Take for instance the gradable adjective *dirty* which we categorized as an absolute adjective previously. Kennedy proposes that this non-vagueness of *dirty* is an inherent characteristic of the adjective brought on by its inherent association with a closed scale structure. We can see that it is indeed a closed-scale adjective with a low scalar endpoint which is used as a minimum standard as it can be paired with diminishers while it cannot be paired with maximizers: *slightly dirty* is a perfectly fine expression, while *absolutely dirty* or *perfectly dirty* is not. Conceptually, we can also confirm that the adjective utilizes a minimum standard: Is a table considered dirty when it has a minimal degree of dirtiness, or only when it has a maximal degree of dirtiness? It is considered dirty if it has any amount of dirt on its surface, therefore a minimal amount of dirtiness. This analysis leads us to conclude that *dirty* is an adjective with a low scalar endpoint which is used as a minimum standard, referred to as a minimum standard adjective.

To take another example, we can look at the adjective *clean* used in (16):

(16)   The table is clean.

Clean is also an absolute adjective as it does not display any of the three signs of vagueness. The adverb test shows that clean is a closed-scale adjective with a high scalar endpoint which is used as a maximum standard: *perfectly clean* is an acceptable expression, while *slightly clean* is not. Conceptually, we can also reason that it is an adjective with a maximum standard: is a table considered clean when it has a minimal degree of cleanliness, or only when it has a maximal degree of cleanliness? A clean table is a table which is free of any dirt: a "less-dirty" table is not a clean table unless it displays a maximal degree of cleanliness. We can analyze *clean*, therefore, as an absolute adjective with a high scalar endpoint which is used as a maximum standard, referred to as a maximum standard adjective. In summary, the crucial connection that Kennedy makes is that endpoints that are embedded in the semantics

of absolute adjectives are used as standards of comparison, with low endpoints as minimum standards and high endpoints as maximum standards. This means that the standard of comparison for absolute adjectives are not derived from a comparison class, as we generally assumed previously, but directly from the endpoints in their scale structures.

On the other hand, relative adjectives show vagueness, and their standard of comparison is derived from a comparison class which can differ depending on context. Kennedy proposes that this is because they naturally carry an open scale, and therefore do not have any endpoints built into their semantics that can function as a standard of comparison. We can take the adjective *fast* as a quick example. As discussed previously, *fast* displays the three signs of vagueness. It is also odd when used with both maximizers and minimizers, as we can see from the oddity of phrases *slightly fast* and *absolutely fast*. Therefore, we can see that it is a gradable adjective which do not have a low or high scalar endpoint, and hence an open scale. We can also see conceptually that *fast* does not have any maximum or minimum standards: it is not that one is fast when it displays a minimum or maximum degree of speed, but only when it displays a degree of speed above a certain standard.[7] Therefore, *fast* does not come with a point which can be used as a standard of comparison built into the semantics of the adjective – it must be derived contextually from a comparison class; and this is the case for all relative adjectives.

To summarize, Kennedy proposes that there is a causal relation between the scale structure of gradable adjectives and the distinction of gradable adjectives into relative and absolute adjectives. The reason why absolute adjectives are not vague is because they are inherently associated with a closed scale structure with endpoints on either side of the scale, which are used as standards of comparison. Because this is part of their semantic meaning, they do not display vagueness no matter under which context they are used. Relative adjectives, on the other hand, are not associated with any inherent endpoints in their scales, and therefore their standards of comparisons must be derived from context, and hence under many circumstances display vagueness.

2.6 The Principle of Interpretive Economy
As a principle to capture this behavior of the two types of gradable adjectives, Kennedy presents the principle of Interpretive Economy, printed below:

> *Interpretive Economy:*
>
> Maximize the contribution of the conventional meaning of the elements of a sentence to the computation of its truth conditions.

An adjective's scale structure is understood as the conventional meaning of the adjective, as it is inherent to its semantics. Therefore, according to this Interpretive Economy, if an

---

[7] One should not mistake this definition as a minimum standard adjective. Although the standard of speed for one to be considered fast may be a minimum 'threshold', it is not the minimum degree of the scale itself such as it was with *dirty* and *clean*.

adjective has an endpoint built into its scale structure, its contribution is maximized by being used as a standard of comparison. This allows the absolute adjective to be not context-dependent when it comes to creating a standard of comparison. However, relative adjectives are forced to be context-dependent when it comes to determining its standard of comparison because there is no conventional meaning tied with the semantics of the adjectives that can be utilized as the standard of comparison.

Although Kennedy's theory and his principle of Interpretive Economy can account for the general behavior of positive forms of gradable adjectives that we observe, it is still not completely comprehensive to be able to account for the exceptions that we sometimes find with their uses. Researchers such as Cruse (1986) and McNally (2011) have provided evidence that it is not always the case that absolute adjectives are non-context dependent. Take (17) and (18) for example:

(17)   The kitchen knife is clean.
(18)   The surgical knife is clean.

Although *clean* is identified as an absolute adjective, we can see that the standard of comparisons used for (17) is different from the one used for (18). If one imagines that they have both a clean kitchen knife and a clean surgical knife, they will agree to both the statements (17) and (18). However, you would not necessarily perform surgery with a kitchen knife as you would with a surgical knife: although a kitchen knife can be clean, it is not usually clean enough to be safe for surgery. For a surgical knife to be considered clean, it must be sterilized. However, this standard is not applied for a kitchen knife to be considered clean. If *clean* as an adjective is a maximum standard absolute adjective as Kennedy proposed, then how could this be the case? If things are considered clean only when the predicate displays the maximum degree of cleanliness on the cleanliness scale, then how can a clean surgical knife display a degree of cleanliness higher than a clean kitchen knife? Compared to the cleanliness of a clean surgical knife, a clean kitchen knife does not display the maximum degree of the relevant concept, which makes *clean* in this case not an absolute adjective by definition. The exact answer to this question is uncertain and the literature provides different theories: one explanation is that this is a matter of imprecision rather than vagueness, which has a pragmatic cause (Leffel, Xiang & Kennedy, 2016) or that there are two types of vagueness, and when it comes to scalar vagueness as we see with gradable adjectives, the granularity intervals of the scale vary depending on context (Sauerland & Stateva, 2011).

Another example of context dependency displayed by a positive form of gradable adjective can be found in (19):

(19)   The wine glass is full.

The social norm for drinking wine is that the wine in the wine glass is never filled past a certain point, which is usually below the mid-point of the glass. Due to this norm, we consider a wine glass to be full when it is filled up to this point. Therefore, when a wine glass

is full, it is not full to the brim as its conventional meaning may suggest: we clearly take the context in which this adjective is used, and set its standard of comparison accordingly. If an absolute adjective such as *full* is not context dependent, why do we understand (19) the way that we do?

The concepts of imprecision and granularity and their relevance to gradable adjectives are discussed in the literature to help explain why we use absolute adjectives the way that we do in sentences such as (17), (18) and (19). Such examples show that it is simply not enough to claim that all absolute adjectives are not context-dependent and that they always derive their standards of comparison from a given endpoint in their scale structure rather than from a comparison class: some absolute adjectives are clearly context-dependent in certain ways when it comes to creating their standards of comparison. Therefore, Kennedy's principle of Interpretive Economy alone is not sufficient to explain the empirical findings we see regarding absolute adjectives. One may find that an interpretation of the Interpretive Economy can still explain this phenomenon: the principle states that it 'maximizes' the contribution of the conventional meaning, not that it relies on it completely. It may be the case that in sentences such as (19), the conventional meaning is 'maximized', but there are other contextual factors in play that prevents the conventional meaning from being utilized. If we were to adopt this interpretation of the principle, then the definition of the principle must be expanded to explain under which contextual circumstances the conventional meaning is not utilized, and how the standard is derived then. The key point is that although the principle of Interpretive Economy is able to explain the general trend of how absolute and relative adjectives are used in most cases, it must be further expanded to cover all empirical behavior we find with absolute adjectives.

2.7 Two types of adjectives: Toledo & Sassoon (T&S)
While Kennedy emphasized the importance of adjectives' scale structures in explaining the nature of the division of relative and absolute adjectives, Toledo and Sassoon (T&S) (2011) proposed another theory that emphasized on the nature of the comparison class of the adjectives to explain the division of the adjectives. They focused on scrutinizing Kennedy's claim that absolute adjectives are not context dependent, and their central question was to see if it was indeed possible for gradable adjectives to have their comparison class determined without any context. We saw that the principle of *Interpretive Economy* presented by Kennedy suggests that for all cases, the scale structure built into the semantics of a gradable adjective always takes precedence over context when we select a comparison class for a gradable adjective. T&S point out flaws that they find in Kennedy's theory, such as the example given above with (17), (18) and (19) and propose that we must look at qualities other than just the scale structure in order to gain a full understanding of how comparison classes are determined for all gradable adjectives.

T&S proposed that the standards of comparison for both absolute and relative adjectives are derived from a comparison class, unlike Kennedy who proposed that it was only the case for relative adjectives. T&S proposed that there are two fundamentally different types when it comes to the nature of a comparison class, and these types are pre-determined for all gradable

17

adjectives. In other words, gradable adjectives can be divided into two distinctly different categories. In this thesis, I refer to their theory as the T&S theory.

The first adjective class is one whose nature of the comparison class is an 'extensional set', meaning that the comparison class consists of different entities or individuals: adjectives that are assigned this comparison class by nature have a comparison class where the different members within the set are unique individuals. I refer to adjectives with this nature as **between-individual** adjectives. The second adjective class is one whose nature of the comparison class is a 'counterpart set', meaning that the comparison class consists of different 'counterparts' of the same individual referred in the argument: adjectives that are assigned this comparison class by nature have a comparison class where the different members within the set are different 'stages' of the same individual that display different degrees of the property relevant to the gradable adjective. I refer to adjectives with this nature as **within-individual** adjectives. I take the problem we encountered with (17) and (18) which are repeated below, as examples to elaborate on the difference between the two types.

(17)   The kitchen knife is clean.
(18)   The surgical knife is clean.

The problem with (17) and (18) as we observed was that although the gradable adjective used in both sentences were identical, the standard for 'maximum degree' was different. By T&S theory, the difference could be explained without having to consult any other concepts such as imprecision – it can be solved just by comparing the nature of the comparison class for the adjective *clean*. According to T&S, *clean* is an adjective whose nature of the comparison class is one of a counterpart set. This makes the adjective a within-individual adjective (how the type is decided for each adjective is explained later in the section). This means that for the adjective *clean*, the members of the comparison class of the subject that the adjective modifies are different stages of itself which display different degrees of cleanliness that the specific knife has displayed in the past. Thus for (17), the comparison class consists of different stages of the same kitchen knife in question, displaying different 'prior degrees' of cleanliness that the knife has displayed in the past. If *clean* happened to be a between-individual adjective, then the comparison class would be different: its comparison class would consist of other kitchen knives. In simpler words, the truth value of statement (17) is dependent on how clean that specific knife usually has been in the past – the standard of cleanliness of the knife is determined as a degree of cleanliness that if the knife displayed a degree of cleanliness above this degree, its cleanliness would stand out as clean compared to the different degrees of cleanliness it usually displays. T&S state that the nature of the scale, whether it is open or closed, is imposed by the individual under consideration. Therefore, the degree of comparison is not a semantic maximum degree based on an arbitrary scale, but the degree displayed by a maximally clean counterpart of this specific knife. As long as this knife meets that standard derived from its within-individual comparison set, it is considered clean.

Once we apply the T&S theory for (18) as well, we can see how the problem we encountered can be explained. A kitchen knife and a surgical knife both have a different standard of

cleanliness because the maximally clean counterpart of the two knives differ due to their function. Because a surgical knife requires a much higher standard of cleanliness due to its role of performing surgery compared to a kitchen knife, the 'prior degree' of cleanliness usually displayed by a surgical knife is relatively high even when it is considered 'dirty'. Therefore, for a surgical knife to be considered clean it must display an absolute degree of cleanliness in order to 'stand out' among its other members of the comparison class (different stages of itself), which all display a relatively high degree of cleanliness. This explains how a kitchen knife and surgical knife can both be clean, and at the same time the surgical knife can be cleaner than the kitchen knife: the within-individual stages of a surgical knife have a higher standard of cleanliness, which speakers know by world knowledge and context, and therefore can be cleaner than a clean kitchen knife.

The matter of how one determines whether a given adjective is within-individual or between-individual is also explained by T&S. The comparison class of an adjective is decided by the variability of the property they evaluate. For example, adjectives such as *clean* or *full* evaluate a property which are manifested at different levels on different occasions. In other words, they represent properties that can vary within a short time period by their typical arguments: the cleanliness of a knife can change quickly, and fullness of a vessel can change quickly. These properties are called **transient**, and adjectives which hold these properties become within-individual adjectives. Other adjectives such as *tall* and *wide* are the opposite, where the property under discussion held by its typical arguments do not vary very often: height of a building or width of a road do not change easily, for instance. These properties are called **enduring**, and adjectives with enduring properties become between-individual adjectives. An important aspect to this division proposed by T&S is that it not only explains some of the flaws present in Kennedy's theory, but can account for all the strengths of Kennedy's proposals also. T&S proposes that within-individual adjectives by nature tend to make their interpretations relative to an endpoint standard, while between-individual adjectives tend to make them relative to a midpoint standard. This is due to a grammaticalization principle they propose which encodes the type of standard usually selected for an adjective as a default convention for these adjectives. Absolute adjectives are typically predicated of individuals whose counterparts have a clear norm-based maximal or minimal elements, meaning that it is common for them to display zero or complete degree of the relevant property in order to be described by an absolute adjective. For example, the adjective *closed* usually predicates over objects like a door, which is considered closed when it displays zero degree of openness – when we consider a door to be closed, the degree of openness it displays is zero, and another door cannot be 'more closed' than this closed door. Therefore, a grammaticalization principle encodes a convention for selecting a maximum or a minimum standard. Relative adjectives are typically predicated of individuals whose distribution of their comparison class is normal and is without any maximal or minimal elements. Therefore, a midpoint standard is encoded as a convention for these adjectives. Examples and details of the grammaticalization principle is provided in 2.3.1 of Toledo and Sassoon (2011).

This leads to the conclusion that within-individual adjectives are what Kennedy calls absolute adjectives, and between-individual adjectives are relative adjectives. A central difference between T&S theory and Kennedy's proposal is that T&S theory suggests that underlying scale structure of adjectives is not the fundamental determiner of the standard of comparison for adjectives. Rather, it is the nature of the property they represent, whether it is enduring or transient. Scale structure only happens to correlate with the two different divisions of adjectives due to how we use these adjectives, but it is not the factor that decides the standard of comparison for the arguments attached to gradable adjectives in their positive forms, as Kennedy proposes.

2.8 Criticisms of the T&S theory
In short, Kennedy proposed that the scale structure of gradable adjectives influenced how comparison classes for those adjectives are established: if adjectives have an open scale (relative adjectives) then the standard of comparison is established through context. If they have a partially or fully closed scale (absolute adjectives), then the endpoints of the closed scale are used as the standard of comparison. T&S proposed that although scale structure of adjectives correlate with their context dependency, they do not fundamentally decide the comparison standard for adjectives. Instead, comparison classes are decided on whether the property they represent are typically transient or enduring: if they are typically transient, then they adopt a within-individual set as their comparison class, and they are also associated with minimum or maximum standards due to the principle of grammaticalization (absolute adjectives). If they are typically enduring, then they adopt a between-individual set as their comparison class, and they are also associated with a midpoint standard (relative adjectives). In this way, both absolute and relative adjectives are context dependent when it comes to building their comparison classes.

Although T&S addresses flaws given by Kennedy, I believe their theory is not fully comprehensive. In this section, I raise several aspects regarding the T&S theory that I believe require further explanation. The first point is that it is unclear how one decides whether an adjective is by nature transient or enduring regarding its property. As mentioned previously, T&S proposes that this depends on their typical arguments. Adjectives exemplified in the paper such as *clean* and *full* have usages that most speakers will agree that corresponds to a transient property. However, there are also adjectives where it is unclear what their typical arguments are – take *opaque* for example. The property of opaqueness (or transparency) is a property that can commonly be transient or enduring depending on its argument, and it is hard to describe which is the 'typical' one. A dirty window which is usually clean can be described as opaque if it is dirty enough to not be fully transparent, which in this case opaqueness would be a transient property, given that a window can easily be cleaned to rid of its opaqueness. However, opaqueness is also a common property to characterize types of rocks – and since the opaqueness of rock types is not something that changes over time, the property of opaqueness would be an enduring property. To summarize, the property of opaqueness is a property whose 'typical' arguments seem to include subjects that can both be transient and enduring with respect to opaqueness. So then, how do we decide what its typical

arguments are in order to decide to assign the adjective its type of comparison class? The T&S theory assumes that all adjectives have clear typical arguments and does not expand on this issue.

The second point follows from the first, in that even if we were to assume that all adjectives do have clear typical arguments such as *clean* and *full*, it does not seem to the case that the type of adjective will stay unchanged when used with atypical arguments. This issue is highly relevant to the current thesis and is a theoretically important issue, because it is an issue directly related to the semantics of the positive forms of gradable adjectives. According to my understanding of the theory proposed by T&S, comparison classes for all adjectives regardless of whether they are relative or absolute, are established after we consider the nature of their property (enduring or transient). This can only be done after considering its usage with its typical arguments, and this is why T&S claims that all gradable adjectives, no matter relative or absolute, are context dependent. However, it is unclear what happens to the comparison class once the arguments that it is being used with are atypical. If T&S are proposing that these comparison classes are built into the semantics of the positive form of the gradable adjectives either in their positive null morpheme POS or in the semantics of the adjective itself, then the comparison class for each adjective should be constant no matter which argument they modify, even if they are atypical arguments. However, this does not seem to be the case, which we can see with the example of *opaque*. Let us assume that somehow, we found a way to explain that a rock is not a typical argument of *opaque*, but a window is, and therefore it took on a transient property as its nature. This would mean that the within-individual comparison class is embedded into the semantics of the adjective, either in the semantics of the adjective or in the slot k in the POS of the adjective. If this was the case, then even when we predicate the adjective over an atypical argument, this comparison class should still apply a within-individual comparison class on its argument. However, this is not what we see with *opaque*. When it is used with respect to rocks, it takes on a between-individual comparison class: whether a rock type is opaque does not depend on the degrees of opaqueness the rock has shown before, but if it relatively opaque or not compared to other rock types. This evidence goes against the idea that the type of the comparison class is predetermined for the adjective and is applied to all arguments. If we were to assume that it is predetermined, then we must also elaborate on when this predetermined type is not utilized.[8]

This phenomenon is not solely restricted to the adjective *opaque,* adjectives with totally closed scales, or with adjectives with unclear typical arguments. We take a look at (17) repeated below, which uses the transient property adjective *clean*:

> (17)   The kitchen knife is clean.

---

[8] Claiming that it is not utilized when it is used with atypical arguments seems to be a weak argument. First, it was unclear how typical and atypical arguments were decided for the adjective. Secondly, if it differs every time it is used with atypical arguments, then there is no empirical merit in claiming that there is a predetermined type in the first place – the same outcome can be observed by claiming that the type of the comparison class is determined by context every time it is used.

As we have discussed previously, *clean* is an adjective whose typical arguments are easily agreed upon relative to adjectives like *opaque:* the typical subjects are ones like *knives*, whose property of cleanliness is high variable between short times. Because *clean* has typical arguments with a transient property of cleanliness, the T&S theory predicts that it is a within-individual adjective. We understand this to mean that this information regarding the adjective is built into the semantics of the adjective. Now take for example (18), which also uses the typically transient adjective *clean,* but used with an atypical argument:

>   (19)   The atmosphere of Neptune is clean.

Let us assume that the atmosphere of an unoccupied planet in space has no reason to vary in cleanliness over time: if there was any adjective whose typical argument was the atmosphere of a planet, then it would be likely that the adjective would be a between-individual adjective, as the argument displays an enduring property. However, if this argument is used with the within-individual adjective *clean*, then what would the comparison class be? If the comparison class is built into the semantics of the adjective, then the comparison class for the atmosphere of Neptune in (19) should be the atmosphere of Neptune in 'prior degrees' of cleanliness that it has displayed in the past, just as it was for a kitchen knife in (17). This would prove to be problematic, however, if the level of cleanliness of the atmosphere of Neptune has stayed exactly the same ever since its creation: how could we derive a degree as a standard of "standing out" for the cleanliness of Neptunes' atmosphere if all its prior degrees are the same degree? If the degree of cleanliness of the atmosphere was the same for millennia, and the cleanliness of the atmosphere at a given point in time is identical to all other points in time, then is the atmosphere neither clean nor dirty? The T&S theory does not make any predictions regarding these kinds of circumstances where the argument of a given adjective is of an atypical type.

Additionally, the intuitive action for many readers to determine if the assertion given in (19) is true or false would be to compare the atmosphere of Neptune to the atmosphere of other planets, such as the one of Earth. In this case, the argument of the within-individual adjective *clean* would be demonstrating a comparison class of whose members are atmospheres of other planets, which is the characteristic of between-individual adjectives. Although this could be explained if (19) had presented an explicit comparison class such as with an additional *for a planet's atmosphere* added to the sentence, if readers indeed take this comparison class for sentences like (19) where there is no explicit comparison class stated, the theory proposed by T&S would falsely predict that (19) would display a within-individual comparison class. Clarification of this issue is extremely important in order to be confident about how much information regarding the standard of comparison and comparison classes are held within the semantics of the positive forms of gradable adjectives.

The third point is that when T&S claim that all adjectives are context dependent, they are only referring to the fact that they are context dependent when deciding whether they are a within-individual or a between-individual adjective, as we must look at context to determine what their typical arguments are. But as we observed with assertion (19) and with the

example of *opaque,* it seems that certain context factors play a role in establishing a comparison class that is different from the one that is 'assigned' to the adjective. Therefore, context may influence our understanding of gradable adjectives in more ways than just the one proposed by T&S. It seems true that all adjectives are context dependent, but it also seems too limiting to state that it is only context dependent for the reason stated by T&S. In order to expand their theory, we must explore in which other ways gradable adjectives are influenced by context – if they are context dependent in one way, it does not seem illogical to assume that they can be context dependent in other ways. If they are indeed context dependent in other ways, then it could expand the theory given by T&S and provide solutions to problems we encountered with their theory such as the one mentioned regarding (19).

To summarize, T&S proposes a theory which not only can account for all of Kennedy's theory but also account for the flaws that Kennedy's theory seem to have. However, some of T&S definitions and proposals are not perfectly clear and does not bring to the table all necessary empirical evidence to support their theoretical claims. Mainly, experimental evidence showing that the comparison classes for gradable adjectives are set in the semantics of positive forms of adjectives through their typical arguments, and that they do not change under usage with atypical arguments are necessary to strengthen their theory.

3. Experiment

The experiment in the current thesis was designed to test the predictions made by the theory proposed by T&S above. Specifically, the experiment is aimed to test the proposal made by T&S that the type of comparison class for all adjectives with a typical transient property is within-individual, while the type of comparison class for all adjectives with a typical enduring property is between-individual. The experiment manipulates the context in which positive forms of different gradable adjectives are used to test whether context of the sentence influences our interpretation of the semantics of positive forms of gradable adjectives by assigning a different comparison class than the one assigned and built into their semantics by their typical arguments, as T&S proposes. Through its result, the experiment can provide additional information about whether the semantics of the positive form of gradable adjective indeed hold a predetermined type of comparison class, and if any additional context factors play a role in the construction of comparison classes for gradable adjectives.

The hypothesis for the experiment is as follows: the construction of comparison classes for gradable adjectives is not strictly dependent on whether an adjective is transient (correlates to absolute adjectives) or enduring (correlates to relative adjectives) in its nature. The context in which an adjective is used can influence the nature of the comparison class for the adjective, for both absolute and relative adjectives.

3.1 Methodology

3.1.1 Material

The experiment was conducted online using Amazon Mechanical Turk. The participants read 22 short stories and each story was followed by a one-sentence declarative statement

regarding the story. Every item used in the experiment is presented in the Appendix. The task of the experiment was to read the statement after reading every story and express on a Likert scale of 1 to 5 (1 being strongly disagree, 5 being strongly agree) how much they agreed with the given statement. Every story was independent, and it was evident that the content of one story was not relevant to the content of another. Every story and its statement were presented separately on their own page. The page moved on to the next page to show a new story only after the participant had entered their answer regarding the previous story, and there was no time limit. The response time for each story was recorded for all participants as well as the value they gave on the Likert scale.

A set for a single participant included 10 experimental materials and 12 filler materials. The fillers were stories whose length and structure were similar to the experimental stimuli. The given task for the filler items was identical to the given task of the experimental items. However, the statement given with the filler items did not contain any positive forms of gradable adjectives, and they had a "correct" answer which the participants were expected to give, which was not the case for the experimental stimuli. This was done to be able to exclude unreliable participants for data analysis. For example, the story of the filler item number 17 in the Appendix explicitly states within the story that Elise wants to make a living as a jazz pianist. The evaluative statement that followed this story was *Elise wants to become a jazz pianist.* For this filler, it was expected that reliable participants would give 5 or at least 4 as their answer. The filler items were divided into three types, where the expected value of their reply was either designed to be a high value (completely agree with the statement), a low value (completely disagree with the statement), or a medium value (uncertain). The fillers with expected high or low values had statements that referred directly to explicit facts introduced in the story, where the answer was clear. For fillers with an expected medium value, the statement was not relevant to the facts presented by the statement but rather the participant's opinion regarding the situation presented in the story. The response time and the response value given for the fillers were recorded in the same way as the experimental material.

The experimental material was a 2 x 2 factorial design with two independent variables with two levels each. The first independent variable is what I refer to as 'adjective-type' in this thesis and is divided into two levels: absolute and relative adjectives. The adjectives used for the absolute level were absolute adjectives which had a clear transient nature respect to their typical arguments – skeptical adjectives such as *opaque* were not used. The adjectives used for the relative level were those considered to be clearly enduring in their nature by their typical arguments. The second independent variable is what I refer to as 'story-type' in this thesis, which is divided into two levels which I refer to as story A and story B. Thus, unlike the filler items, every story related to an experimental item had two different versions of the story. Every participant only received one level of the story, either story A or B, but not both, per adjective.

3.1.2 The difference between story-type levels (story A and B)
The difference between story A and B of the variable story-type is the difference in the

variability of the property of the adjective. Figure 2 below clarifies the design for absolute adjectives with maximum standards.
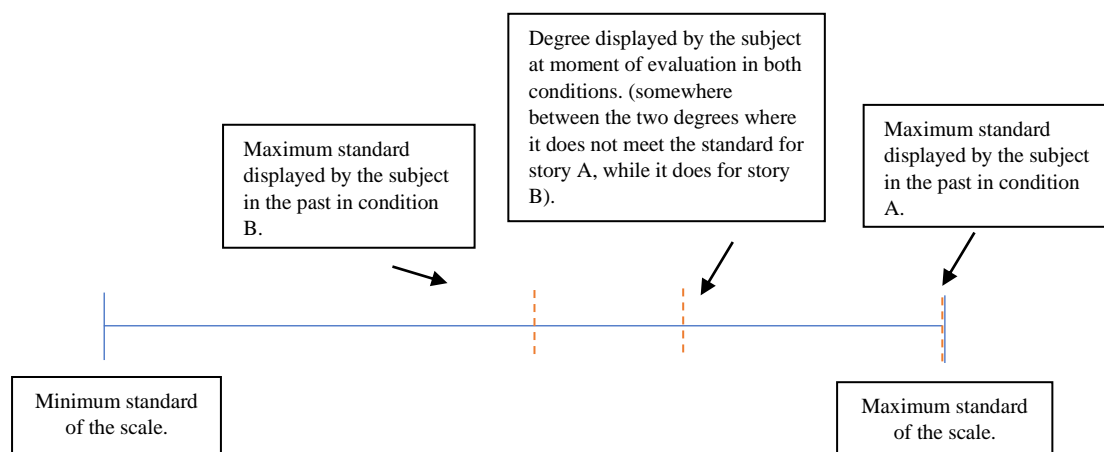

Degree displayed by the subject at moment of evaluation in both conditions. (somewhere between the two degrees where it does not meet the standard for story A, while it does for story B).

Maximum standard displayed by the subject in the past in condition B.

Maximum standard displayed by the subject in the past in condition A.

Minimum standard of the scale.

Maximum standard of the scale.

Figure 2. Scale representing the general difference between story A and story B used for absolute adjectives with maximum standards.

Story A provides a context where the prior degrees of the property relevant to the testing adjective usually displayed by the subject is much closer to the scalar endpoint of the adjective. In other words, the subject in story A usually displays a "higher" degree of the property. This standard is a maximum one if the adjective is a maximum standard one (or a totally closed scale which utilizes its upper scalar end point), and the standard is a minimum one if the adjective is a minimum standard one (or a totally closed scale which utilizes its lower scalar end point). This standard is shown as the rightmost dotted line in Figure 2.

Take for example item number 1 in the Appendix, which shows the experimental material for the absolute adjective *clean* used in the experiment. All participants received either story A or story B as one of the 22 stories they had to read during the experiment. Half of the participants received story A, while the other half received story B. In story A, a surgeon named Dr. Janssen takes home a surgical knife that he has been using for 2 years in the hospital for surgery. He intends to use it as a fruit knife at home. After taking it home, he washes it in the dishwasher, along with all his other kitchen knives. Story B is identical to story A, with the only difference being that the surgical knife he has taken home was never used in the hospital. He took a brand-new surgical knife from the hospital, and he has been using it as a kitchen knife at home for the past 2 years. We can see that the maximum degree of cleanliness displayed by the surgical knife in story A is much higher than the maximum degree of cleanliness displayed by the surgical knife in story B. Whether the participant read story A or story B, the statement that they had to evaluate after reading the was identical for both stories: *After being washed in the dishwasher, Dr. Janssen's knife is clean*. We can see that for story A, the cleanliness of the knife after being washed in the dishwasher **does not**

meet the maximum standard of cleanliness that the knife usually displays in the story – the visualization of degrees displayed in the story regarding *clean* is presented in Figure 3:
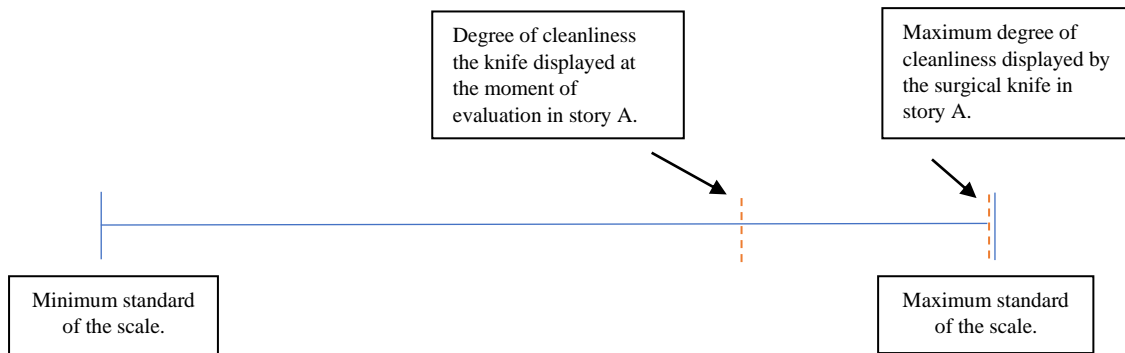

Figure 3. Visualization of degrees of cleanliness displayed by the surgical knife in story A.

This is different from how story B is designed. For story B, the cleanliness of the knife after being washed in the dishwasher **does** meet the maximum standard of cleanliness that the knife usually displays in the story – the degree of cleanliness it displays at the moment of evaluation is not higher[9], but exactly the same as the maximum degree of cleanliness that it usually displays: the visualization of story B for the adjective *clean* is presented in Figure 4:
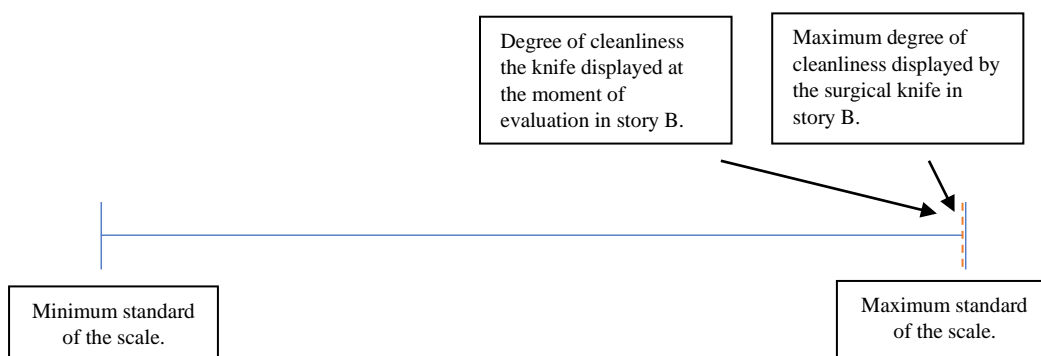

Figure 4: Visualization of degrees of cleanliness displayed by the surgical knife in story B.

---

[9] Story B for other adjectives, such as *dirty*, has the degree of the property at the moment of evaluation exceed the usual degree of the property displayed by the subject. However, this was not expected to be a significant difference as the requirement for the story was that it meets the usual standard. Whether it just meets the standard, or exceeds it, theoretically it makes no difference in the design.

The stories A and B were designed to test the T&S theory in the following way: According to the T&S theory, *clean* is a gradable adjective that denote a transient property with respect to their typical arguments. Therefore, it is assigned a within-individual comparison class. This means that when we evaluate assertions with positive forms of the adjective *clean* used with their arguments, such as *the knife is clean*, we decide whether the knife is clean by seeing if the level of cleanliness it shows meets the standard of cleanliness it usually shows. Under this theory, the knife under evaluation in story A should not be considered clean (producing a low response value to the statement), while it should be considered clean for story B (producing a high response value to the statement). The material for absolute adjectives with minimum standards, such as *dirty*, was also designed to create the same outcome: the subject in story A at moment of evaluation should not be considered to display that property, while it should in story B. The general design for absolute adjectives with minimum standards is visualized in Figure 5 below:
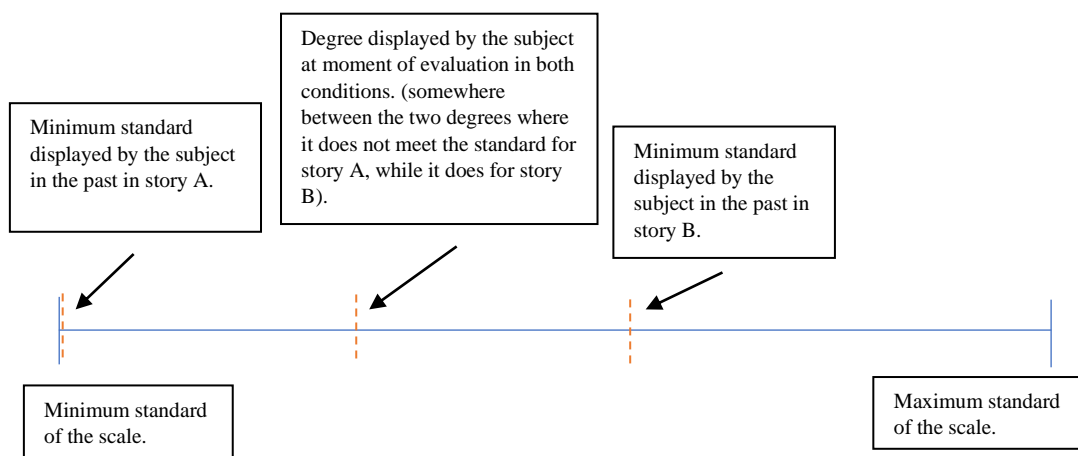

Figure 5. Scale representing the general difference between story A and story B used for absolute adjectives with minimum standards.

The same design was applied to relative adjectives, and the general design for relative adjectives is visualized in Figure 6 below:
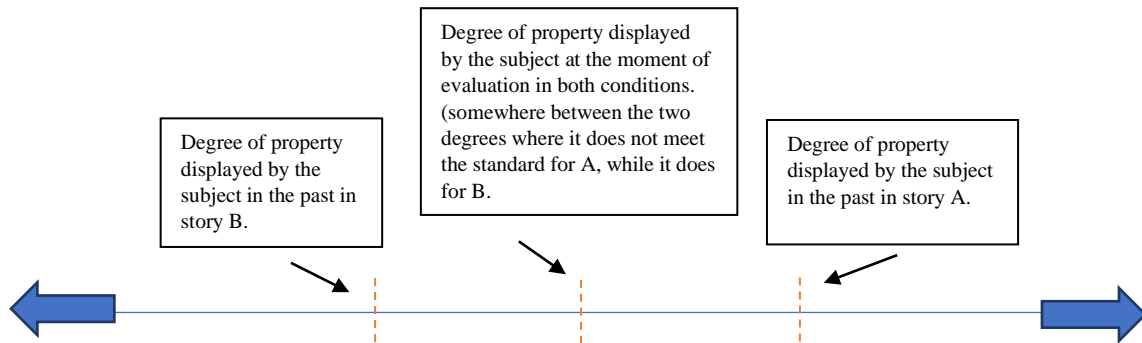
Degree of property displayed by the subject in the past in story B.

Degree of property displayed by the subject at the moment of evaluation in both conditions. (somewhere between the two degrees where it does not meet the standard for A, while it does for B.

Degree of property displayed by the subject in the past in story A.

Figure 6. Scale representing the general difference between story A and story B used for relative adjectives.

The design of the stories is the same for relative adjectives, with the only difference being that since relative adjectives do not naturally make their interpretations relative to an endpoint, the degree of property displayed by the subject in the past in condition A is not relative to an endpoint. However, the structure of the design is similar to the structure for absolute adjectives, in the sense that the usual degree of the property displayed by the subject in story A is high, while it is lower for the subject in story B, and the degree displayed by the subject at the moment of evaluation is a degree that is higher than the usual degree displayed by the subject in story B but lower than the usual degree displayed by the subject in story A.

3.1.3 Prediction
Under the assumption that the theory proposed by T&S is correct, we can make two distinct predictions regarding the outcome of the experiment. The first prediction is a prediction regarding absolute adjectives, mentioned above: for absolute adjectives which are assumed to carry a within-individual comparison class, there will be a difference between how the participants react to the given statement depending on whether they receive story A or story B. In other words, the independent variable of story-type should cause a difference between the response values of statements for absolute adjectives. The second prediction is that the independent variable of story-type should not cause a difference between the response values of statements for relative adjectives. This prediction is made by the T&S theory because the context difference between story A and B is with respect to the degree of standard displayed by a specific individual, while the comparison class for relative adjectives is assumed to be between-individual by nature. Because their comparison classes are between-individual by nature, even with the same contextual manipulation as there was with stories related to absolute adjectives, it is predicted not to influence the standard of comparison for relative adjectives because their comparison class is not composed of differing degrees of the property displayed by different stages of the specific individual. To make sure that the difference between story type A and B only includes a difference regarding the degree of

relevant property displayed by the specific subject of the story, the stories were made to be minimal pairs which differ only by one word or differ in fewest words possible. It is true that even with a minimal difference, the two stories ultimately provide two different contexts. Therefore, it is possible that the difference between story A and B will influence the response value of relative adjectives in an unforeseen way. Even if this would be the case, however, I predict that the response value will ultimately not be different between story type A and B for relative adjectives, because it would then be a confounding factor that exists in no particular direction throughout all versions of all the stories used for testing. Because the effect of this potential confounding factor exists in all directions, cumulatively it would make no difference in the overall response value of each adjective.

3.1.4 Participants

Data from a total number of 70 participants were collected, who were paid $1.3 to participate in the experiment. 24 out of the 70 were excluded from data analysis for the following reasons: one participant indicated that their native language was not English. One participant had their data deleted for an unknown reason. 10 participants had an average response time of 6 seconds or faster for each story.[10] These participants were excluded from analysis as the response times made it clear that they did not read the given stories carefully, and therefore their answers could not be reliable. 12 participants had given values for the filler items in a way that significantly differed from what was expected, and their response times were also relatively fast, between 6 and 10 seconds. These participants were also excluded as it was evident from their answers to the filler items that they did not read the given stories carefully.

After 24 participants were excluded, 46 participants were left for data analysis. There was an equal number of participants who received version A and B for each experimental item, meaning that for every experimental item 23 participants received version A and the other 23 participants received version B.

3.2 Results

A two-way ANOVA with interaction was performed using R to test for an interaction effect between story-type and adjective-type on the response value, as well as the main effect of each independent variable on the response value. A significant main effect of story-type on response value was found, $F(1, 456) = 4.00$, $p = 0.05$, and a significant main effect of adjective-type on response value was also found, $F(1, 456) = 14.69$, $p < 0.01$. However, no significant interaction effect was found between story-type and adjective-type on the response value, $F(1, 456) = 0.67$, $p = 0.41$. Boxplots showing the response value for all experimental items are shown in Figure 7 and Figure 8 below.

---

[10] The individual response times for each item was also looked at for these participants. Majority of them had a response time that worked as an outlier that increased the average response time. Most of the response times for these participants were between 1 and 3 seconds.
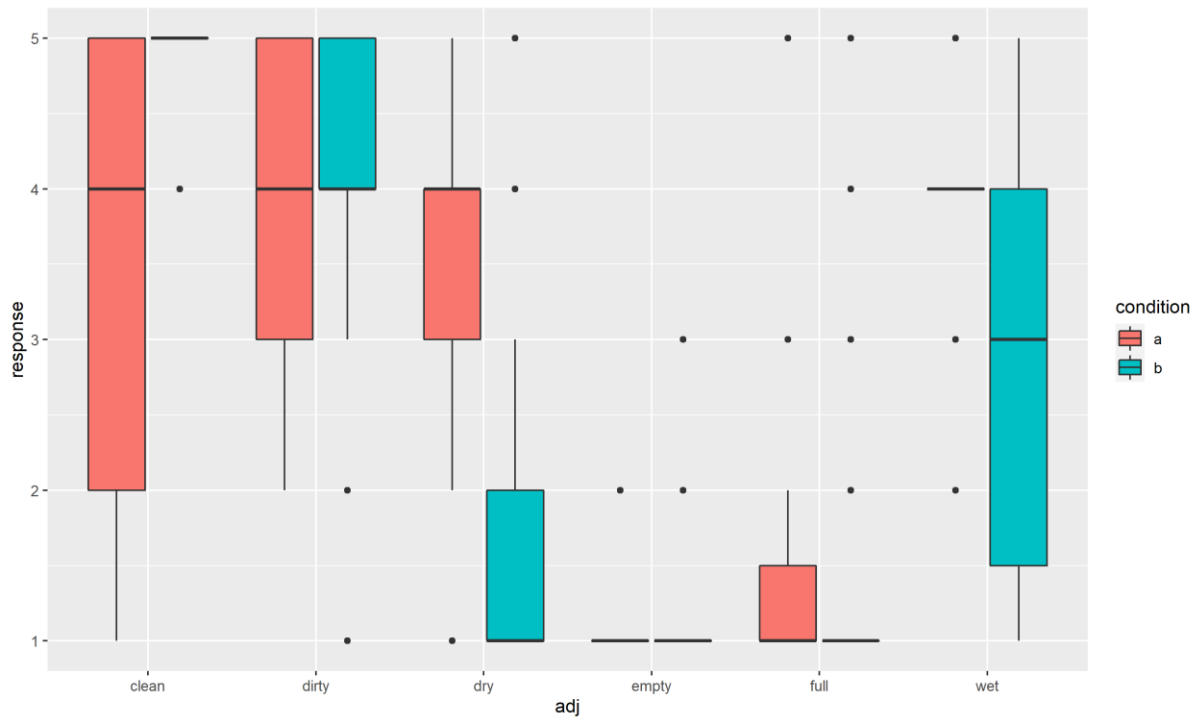
Figure 7. Boxplot showing the response values for each tested absolute adjective.
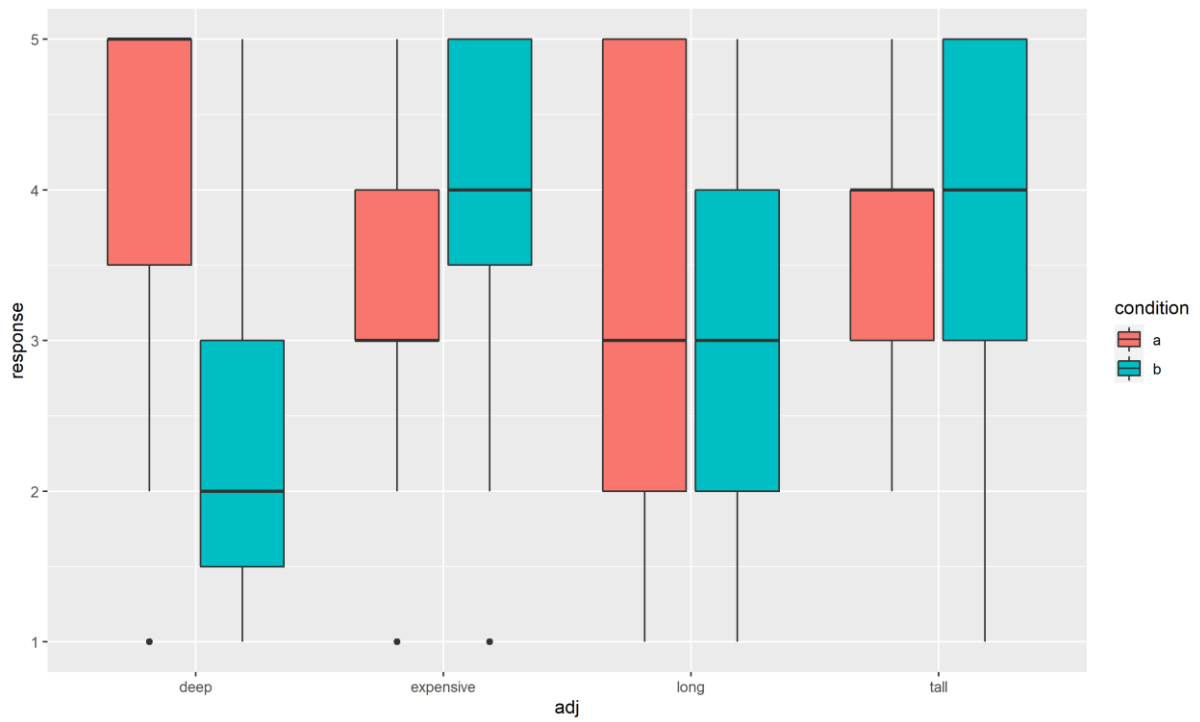

Figure 8. Boxplot showing the response values for each tested relative adjective.

4. Discussion
4.1 Interpreting the results
The results revealed that this experiment did find a significant main effect of both the independent variables of story-type and adjective-type on the dependent variable of response value. However, the experiment did not find a significant interaction effect between the two independent variables. The significant main effect of adjective type on the response value is unsurprising. Absolute and relative distinction within gradable adjectives have a strong empirical evidence in the literature, and how we interpret their meaning is inherently different. Therefore, this result met the expectation that there would be a significant difference on the response value depending on what the adjective type was.

The significant main effect of story-type on the response value helps us to answer the research question, but only partially. This result tells us that the prior degrees (the degree of relevant property usually displayed by the subject in the past) displayed by the subject of the story influenced the response value in a significant way. Since the manipulation of prior degrees of subjects is manipulation of degree displayed by an individual, this result seems to support the claim by T&S that comparison classes for certain adjectives are within-individual. However, the absence of a significant interaction effect between adjective-type and story-type tells us that the manipulation of prior degrees of an individual influences our interpretation of gradable adjectives regardless of the adjective's type, in contrast to what T&S's theory might predict.

One critical assumption, however, in using ANOVA tests to analyze the results of this experiment is that the Likert scale used to score the response value is a continuous variable. This is not fully appropriate since a Likert scale is ultimately a categorical variable. Therefore, I approach the statistical significance shown by the ANOVA test with caution: although the test is harmless in spotting the general trend of the response values, I do not take its statistical significance at face value. In order to analyze the results of the experiment more accurately, rather than analyzing the statistical significance shown by the ANOVA test, I take a qualitative look at the results and see how each material can provide a clue to answering the main research question. In the subsections below, I take specific examples of the material used in the experiment and discuss what their results suggest.

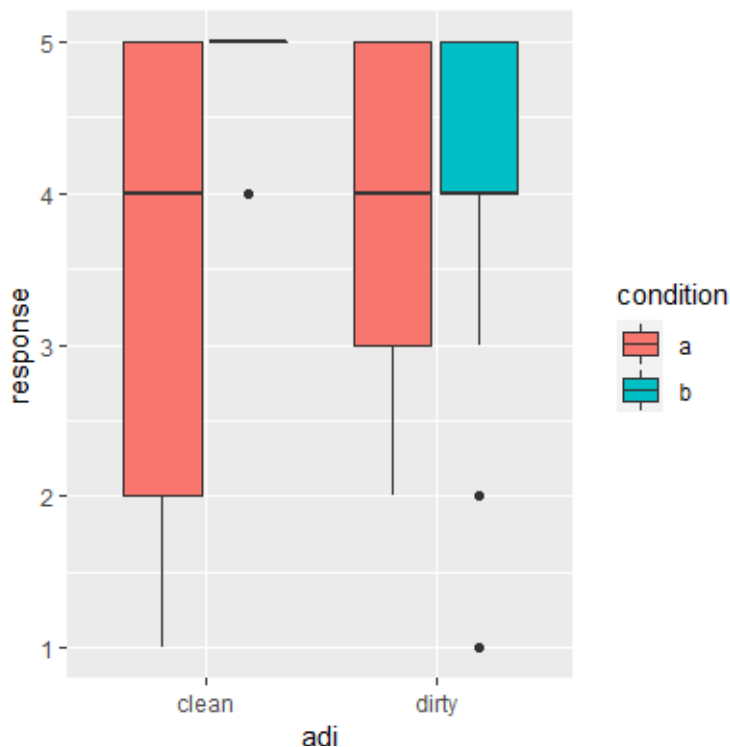**4.1.1 Absolute adjective pattern 1: *Clean* and *dirty***

Figure 7.1. Boxplot showing the response value for *clean* and *dirty.*

The boxplots for *clean* and *dirty* in Figure 7 (shown separately in Figure 7.1) show that the positive forms of these adjectives behave in a way that is most similar to what T&S's theory would predict regarding the behavior of absolute adjectives. Take the results of *clean* for example: the participants were less likely to accept a knife to be clean when the knife's cleanliness level was at "kitchen-level clean" if the knife usually displayed a much higher degree of cleanliness. The participants were more likely to accept the knife to be clean at the same level of cleanliness if the prior degrees (of cleanliness) of the knife was identical to the degree of cleanliness that the knife displayed at the moment of evaluation. If we were to observe the results of *clean* and *dirty* on its own, it would be strong evidence to support the T&S theory: the typical argument for absolute adjectives is transient in nature, and therefore their comparison classes are within-individual. Therefore, the variable of story-type, which manipulates degree in a within-individual manner, creates a completely different set of comparison classes for the knife depending on which version of the story the participants read. Although the pattern observed with the result of these two adjectives seem to strongly support the T&S theory, the other four absolute adjectives tested seem to display a different pattern in their results. Within the class of absolute adjectives, there seems to be two other major patterns that we observe.

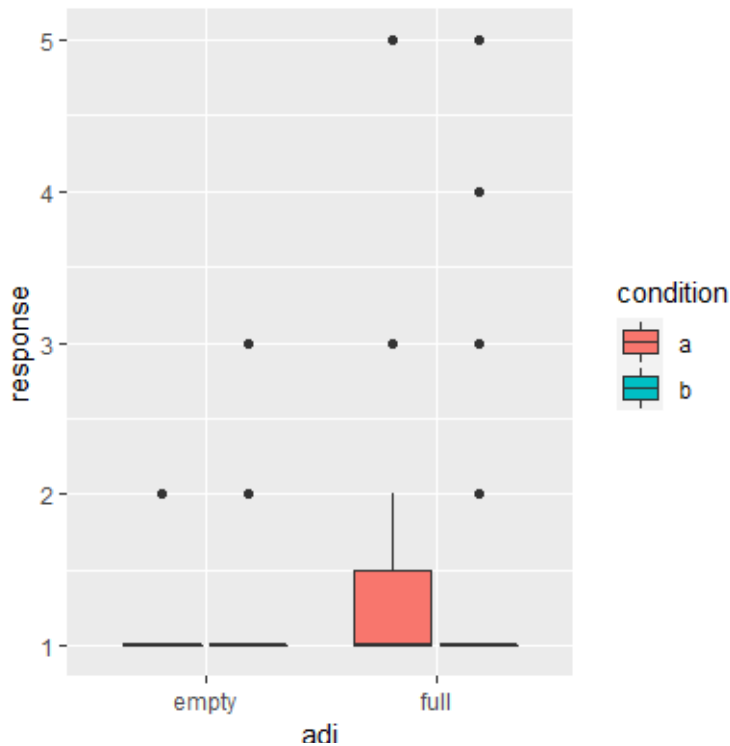**4.1.2 Absolute adjective pattern 2: *empty* and *full***

Figure 7.2. Boxplot showing the response value for *empty* and *full.*

The results from the two adjectives *empty* and *full* are different from what we observed with *clean* and *dirty*: if the results from the material regarding *clean* and *dirty* can be accurately predicted by the T&S theory, then the results from *empty,* and *full* can be accurately predicted by Kennedy's theory. For these two adjectives, the response value does not seem to significantly differ whether the participants received story A or story B. The behavior of these three adjectives is what Kennedy's principle of *Interpretive Economy* would predict: these absolute adjectives hold a minimum or maximum standard within their scale structure as part of their conventional meaning that are utilized for their interpretation, and therefore there is no need to be context dependent to interpret its meaning. For example, no matter what the prior degrees of fullness[11] usually displayed by the gas tank was, the participants refused to refer to the gas tank as *empty* when it had any amount of gas inside the tank. According to the *Interpretive Economy* this is because *empty* inherently has a low endpoint on its semantic scale structure which is used as a minimum standard that is utilized for its interpretation, and thus whether ones' fullness meets this minimum standard or not is the only quality that is considered to determine whether something is empty or not. Since the difference between story A and story B is purely contextual and the scale structure of the adjective remains the same, there is no difference regarding the response value between the two versions. The results were similar with *full,* where a similar reasoning utilizing a maximum scalar endpoint as a standard can be used to explain its results. This result ultimately suggests that T&S's theory does not fully predict the behavior of all absolute adjectives correctly, and that the principle of *Interpretive Economy* suggested by Kennedy can help explain some of the behavior that cannot be predicted solely by T&S's theory.

---

[11] I use fullness here instead of emptiness because fullness is more common to measure the property rather than emptiness – the property they represent is inherently the same, so it does not make any theoretical difference concept is used.

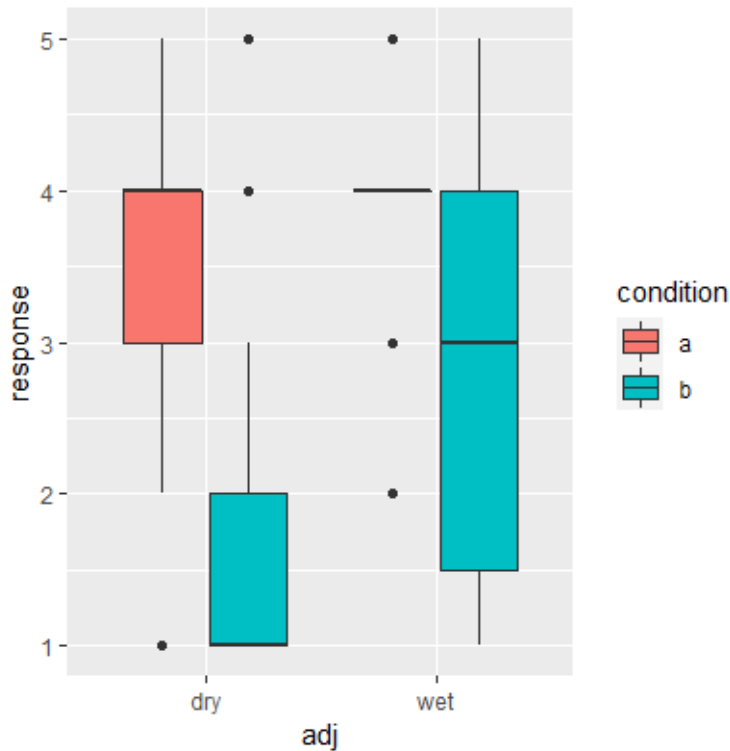### 4.1.3 Absolute adjective pattern 3: *Dry* and *Wet*

Figure 7.3. Boxplot showing the response value for *dry* and *wet*.

Among absolute adjectives, there is yet another pattern that we see from the results that differ from the two patterns we observed above. Unlike the results of *empty* and *full,* the variable of story-type seems to have an impact on the response value for the two antonyms *dry* and *wet*. However, it also differs from the results from *clean* and *dirty* in that the direction of the effect we observe is the opposite. A desert was more likely to be considered dry if the desert had displayed a significantly higher prior degrees of dryness, compared to a desert that had displayed a significantly lower prior degrees of dryness. The direction of this effect is quite the opposite of what we observed with *clean* or *dirty*, and therefore opposite of what T&S's theory would predict. Why were participants likely to agree to a statement referring to the desert as a 'dry desert' and a jungle as a 'wet jungle' if the degrees of dryness and wetness that they displayed, respectively, are significantly lower than the degree of the property that they usually displayed (story A)? And why were participants likely to disagree to the statement even if they displayed a higher degree of that property compared to the degree they usually displayed (story B)? This behavior cannot be predicted by T&S's theory, nor by Kennedy's *Interpretive Economy*. Before I provide the possible ways to account for the differences in the results we find within among absolute adjective types, we first look at the two major patterns we see regarding the results we find with relative adjectives tested in this experiment.

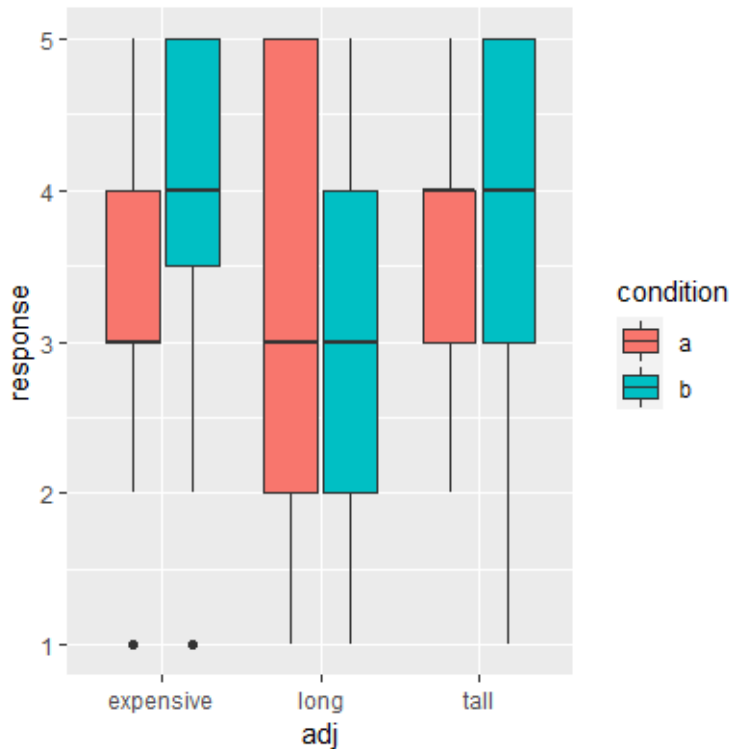**4.1.4 Relative adjective pattern 1:** *expensive, long,* **and** *tall*

Figure 8.1. Boxplot showing the response value for *expensive, long,* and *tall*.

The pattern of the results from the three relative adjectives *expensive, long,* and *tall* are different from any of the patterns we observed with the results from any absolute adjectives. It is most similar to what we saw from the results of *empty* and *full* in the sense that there seems to be no impact of the condition variable on the response value. However, it also differs from the results of the two absolute adjectives in the critical aspect that the median of the results are not value 1 or 5 on the Likert scale, but rather in the middle of the scale, at 3 and 4. This difference is critical as it suggests that the membership of these properties are not as clear as it is with absolute adjectives. The fact that the median was 1 for the absolute adjectives *empty* and *full* suggests that it is very clear when something is full or empty – it shows that people are very certain when a gas tank or a mug is full or empty. The fact that the median is 3 and 4 for *expensive, long,* and *tall* can be used to argue that the membership of these adjectives is not as clear as absolute adjectives: people were in moderate agreement of the given statements, but they were not completely certain in their agreement. This suggests that the interpretation of these relative adjectives is not only not context dependent in a within-individual manner, but they also do not utilize a clear standard for its interpretation either: if it was perfectly clear whether a novel was long or not, for example, it seems odd that the median response value of the statement was a 3 when the statement stated that the novel is long. If it was perfectly clear at which point a novel is long, then the median would have more likely been a 1 or a 5 – as there would be a clear requirement, and therefore people would either strongly agree that the novel is long with a response value of 5, or strongly disagree that the novel is not long with a response value of 1.

This behavior of these three relative adjectives is what T&S's theory predicts: relative adjectives are context dependent, but the members composing the comparison class set are different individuals. Therefore, the comparison class for these adjectives will change if the context manipulates the members of between-individual comparison set. Since the story-type

variable only manipulates the context in a within-individual manner, the context difference provided by the variable has no impact on the response value of these adjectives. Additionally, because they also do not have a standard in their conventional meaning, the response value is also not centered around the end of a given scale.
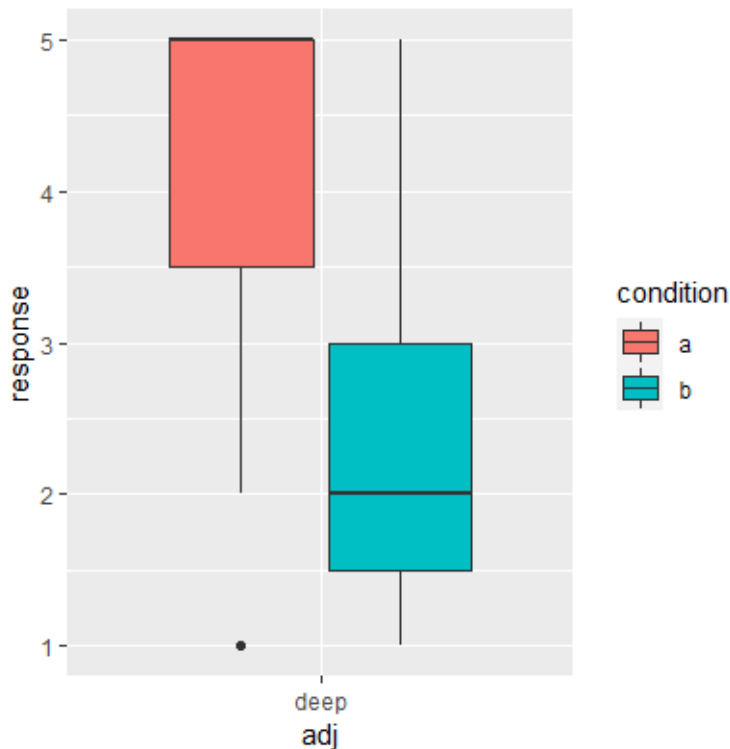
4.1.5 Relative adjective: *Deep*

Figure 8.2. Boxplot showing the response value for *deep*.

The pattern shown by the results related to the relative adjective *deep* is different from the pattern we see in the other relative adjective tested. In fact, it strongly resembles one of the patterns we observed among the absolute adjectives: *dry* and *wet*. As it was for the absolute adjectives, this pattern shown by *deep* is also difficult to explain with either Kennedy or T&S's theories. There are two aspects of this result that seem to be unexplained by Kennedy or T&S's theories. The first aspect is that *deep* is a relative adjective but there seems to be an influence of a within-individual context on the response value. According to T&S, the prior degrees of depth displayed by this well should not influence the decision of whether this specific well is deep or not, because the comparison set is a between-individual one. Both versions of the story told that the well was fixed to have the same depth as other wells that James dug over the year. If this is indeed the case, then it is hard to explain why the response value for story A is so high while the response value for condition B is so low. The second difficult aspect is the direction of the difference we see between the story-type conditions, for the same reason why the direction of the difference we observed with *dry* and *wet* was unexplainable by the two theories: even if we were to assume that story-type could impact its comparison class, it is hard to explain why the well is considered more deep if the well had displayed a higher degree of depth in the past. Table 1 below summarizes whether the patterns we observed in this experiment can be explained by either the T&S theory or Kennedy's principle of *Interpretive Economy*. In the sections that follow, I suggest possible explanations for the patterns we saw that cannot be explained by the two theories, and possible directions of new research that could test the explanations.

Adjective Type	Adjectives	Predicted by either Kennedy or Toledo & Sassoon
Absolute	*Clean*	YES
	Empty, Full, Dirty	YES
	Dry, Wet	NO
Relative	*Expensive, Long, Tall*	YES
	Deep	NO

Table 1. Table summarizing the major patterns we see with absolute and relative adjectives tested.

4.2 Explaining the results of *deep:* The Ship of Theseus paradox

We first start with a possible explanation for the odd behavior shown by the relative adjective *deep.* I propose that this finding is because the property relevant to the adjective, depth, in the scenario given, modifies not only the property of the subject but rather the identity of it as well. This idea is similar to the well-known paradox called the Ship of Theseus paradox: if a ship completely made of wood has its parts changed little by little over the years, and at one point the ship contains none of the original parts it was built with, is the ship the same ship as it was when it was built? This metaphysical question regarding identity can be applied in a similar manner to the story that was used in the experiment for *deep*: if the depth of a well was changed over time, is that well the same well as it was before its' depth was changed? Many of the participants seem to believe that it is indeed not the same well. The story for this adjective told a story where James built dozens of wells over the period of a year for volunteer work, but the very first well that he built could not be used because of its depth. In story A the well was too deep compared to other wells, and in story B the well was too shallow compared to the other wells. At the end of the story in both versions, however, the depth of the well was fixed to be the standard depth just like any other well James had ever dug. The statement that the participants had to rate for agreement was *James' first well is deep*. The intention of the statement was for participants to evaluate the depth of the well after the depth was changed. The phrase *first well* was used to refer to that specific well, because the story also mentioned the existence of several other wells that James had built. However, the participants seemed to have understood the change in depth of the well as a change in the identity of the well. In other words, it seems that the well that was fixed was no longer considered to be the same well as James' first well. So, when the statement refers to "James' first well", they associated this statement with the well before the depth was fixed, but not after. This assumption not only explains the difference between the response value of story A and B but the direction of the difference as well. As the well was perceived to be a different well and no longer "James' first well" once its depth was fixed, the material failed

to introduce a within-individual comparison class to the subject. The well after its depth was fixed became another individual in the between-individual comparison set of the 'first well', which is the well with its original depth. So, when the participants evaluated the statement *James' first well is deep*, they strongly agreed that it was so for story A, which presented a high prior degree of depth, but not for story B.

If modification of some properties relevant to relative adjectives can ultimately modify the identity of its argument, then we can conclude that relative adjectives are context dependent beyond the level of just being context dependent for the purpose of being assigned a comparison class type from its typical arguments: pragmatic context at the utterance level can also influence the composition of its comparison class. This assumption is then also in line with the theory proposed by T&S: ultimately, the comparison class for the adjective *deep* is still between-individual. The difference found between the two-story types for this adjective is not due to having a within-individual comparison class, but rather the context of the story changing the identity of the subject rather just the property of it.

It is not possible to conclude in this paper, however, modification of which property can lead to the modification of identity as this seems to be a question more relevant to metaphysics rather than linguistics. However, there are intuitive differences between certain relative adjectives that can be seen regarding rather it can modify the subject's identity. Take (20) and (21) for example:

(20)   The depth of the well is changed – now it's a completely different well!
(21)   ?The price of the couch is changed – now it's a completely different couch!

It seems acceptable to claim that the identity of the well has changed once its depth has changed, as the felicity of the utterance in (20) shows. However, the oddness of (21) suggests that a change in the price of the item does not have the power to change the identity of the item – it merely has the power to change a property related to it. It seems odd to claim that the couch is a different couch once its price has changed. The difference between the adjective *deep* and *expensive* seems to be that *deep* is relevant to a property that can be critical to one's identity, while *expensive* is not. However, it is important to note that whether a change in the property can ultimately lead to a change in identity seems not to be embedded in the adjective but is somehow context specific – the property can modify the subject's identity, but not always. Take (22) as an example:

(22)   The length of the novel was changed – now it's a completely different story!
(23)   The length of the novel was changed – but it's still the same exact story!

The felicity of (22) suggests that the length of a subject potentially holds the power to influence the identity of the subject, but the felicity of (23) also suggests that it is not necessarily the case. The results of *long* in the experiment tells us that in the context of the experiment, the identity of the novel did not change as a result of a change in its length.

To summarize, the results of the relative adjectives correspond to Toledo & Sassoon's claim that the comparison class of relative adjectives are between-individual. The relative

adjectives seem to be context dependent in the way that they depend on their typical arguments to determine the type of their comparison class, but they also seem to be context dependent at the utterance level in specifying the members of their between-individual comparison set. Specifically, it seems that certain relative adjectives have influence over the identity of the subject that they modify. The details of what makes an adjective identity-related, and under which circumstances the modification of an identity-related property actually allows the modification of the identity of the subject, is unclear; but what seems to be clear is the adjective *deep* and its involvement with modification of identity beyond the modification of property. With this unknown aspect of *deep* unaccounted for, it is best to disregard the pattern we see with *deep* when making conclusions related to the experimental hypothesis.

4.3 Explaining the results of *dry* and *wet*

The pattern we saw with the results with *deep* were similar to the results we observed with the absolute adjectives *dry* and *wet*. However, the same explanation cannot be used to explain the behavior of these two adjectives. We take the results of *dry* as an example. In the story for *dry*, there is a desert whose climate has changed in recent years due to climate change. In story A, the desert is extremely dry (maximum degree of dryness), but in recent years it has become less dry compared to its past. In story B, the desert was not so dry, but in recent years it has become drier compared to its past. The ultimate dryness of the desert of the two versions is the same degree of dryness, where it rains once a month on average. The statement that the participants had to evaluate was *Gerel visited a dry desert*. It was clear from the story that Gerel had visited the desert after the climate around the desert was changed.

We could attempt to explain this result the same way we did with the results of *deep*, but that would be problematic as unlike the example with wells and the adjective *deep,* the statement for these adjectives clearly states that Gerel had visited the desert/jungle after the climate around the desert/jungle was changed. Therefore, even if we did claim that the change in the climate of the desert/jungle changed its' identity in this context, no matter if the comparison class of these adjectives is by nature within-individual or between-individual, the response value for story A should be lower than the response value for story B. Therefore, finding such a pattern within the absolute adjective group is puzzling. One explanation could be that the design of the material used for the theses two stories was flawed and failed to test what it was intended to test. To statement used to describe the climate change that occurred in this desert (and the jungle for *wet*) was "However, due to climate change the weather **around** the desert/jungle has started to change." It is possible that the participants understood the preposition **around** to be the climate not including the desert/jungle itself – so when the climate had changed **around** the desert/jungle, the climate of the desert/jungle was viewed to be unaffected. Therefore, the degree of the desert/jungle intended to be the prior degree was viewed to be the only degree displayed by the desert/jungle, and the participants evaluated the statements according to this interpretation. Whether the results of these adjectives were a product of a mistake in the design can only be concluded after a rerun of a revised testing

material, which there was no time to do so for the current thesis. With this possibility that the pattern we observe with these adjectives is the result of a flaw in the material set-up, unfortunately the results from these adjectives cannot help us to derive any concrete conclusions.

4.4 Explaining the dual presence of Kennedy and T&S in absolute adjectives

The odd results displayed by *deep* was still able to be explained within the boundaries of the theory presented by T&S: all four relative adjectives seem to hold a between-individual comparison class due to their typical enduring nature, and therefore its results are in line with the predictions made by the T&S theory. However, with the odd results of *dry* and *wet* set aside, the results shown by some absolute adjectives are not in line with T&S's predictions. The results of *empty,* for example, clearly shows that the standard of comparison derived from absolute adjectives are not always from a within-individual standard. In the story for *empty,* Emma drives the same car for 5 years. In story A, the gas tank is always filled as little as possible, while in condition B the tank is always filled to the maximum. At the moment of evaluation of the given statement, the gas tank is filled to the half-point of the tank. According to T&S's predictions, the gas tank filled to the half point of the tank in story B should be considered empty as in the past it was always filled to the top, and therefore stands out to in its emptiness compared to its prior degree of emptiness. However, this is not the case, and the response value for this condition was no different from the response value of story A. Similar results were found with *full*, where the variable of story-type that manipulated the within-individual standard of the subject did not seem to have an influence on the response values.

The results found with *clean* and *dirty* seem to be in line with the predictions made by T&S, but the results found with *full* and *empty* critically refute the claim of the T&S theory that absolute adjectives take on an inherent within-individual comparison class due to their typical arguments: the results here seem to suggest that the standard is derived from an endpoint in their semantic scale, as Kennedy's *Interpretive Economy* presents. Whatever the explanation between the two opposing results may be, one conclusion that can be made from the existence of both patterns is that the construction of the standard of comparison seems to be highly context dependent at the utterance level. Therefore, neither Kennedy's claim that absolute adjectives are not context dependent, nor T&S's claim that absolute adjectives hold inherently a within-individual comparison class can fully explain the results of this experiment: it is clear that the construction of the standard of comparison is highly variable by context.

4.5 An alternative proposal: a weaker interpretation of the T&S theory

The overall purpose of this experimental thesis was to test the predictions made by Kennedy and T&S, as well as expand their theories in detail if necessary. The results of the experiment have put to test the predictions made by the two theories and met with results that seem to suggest that although both theories seem to be partially true, they are not independently comprehensive enough to be able to explain all empirical results we have found through this experiment. It seems to be the case that the process of deriving a standard of comparison for

all adjectives is not purely dependent on their semantic scale nor on their typical arguments. Rather, it seems that the process is highly context dependent. In this section, I propose a theory that combines the two ideas presented by Kennedy and T&S in order to present a theory that can best account for the results we have found in this experiment.

First, the results of this experiment seem to agree with Kennedy's proposal that the distinction between relative and absolute gradable adjectives comes from the structure of their semantic scales. The results of relative adjectives showed that none of the adjectives were interpreted with respect to an endpoint of a scale (derived from the finding that the median response values for these adjectives are not 1 or 5). I analyze that this is simply because they do not have an endpoint in their scale, as Kennedy proposes. On the other hand, the results of absolute adjectives tested showed that they were interpreted with respect to an endpoint of a scale. However, while absolute adjectives showed that this endpoint-evaluation tendency was possible for them, it was not a necessity: results of *clean* and *dirty* showed that their interpretation can also be influenced by contextual factors, specifically by manipulation of maximum or minimum degree of standard displayed by the individual in the past. In this way, this thesis presents evidence against Kennedy's proposal that absolute gradable adjectives always derive their standard of comparison from their scalar endpoints and are not context dependent. The T&S theory seems to be able to account for the finding that Kennedy could not account for: it seems to correctly predict the results of relative adjectives, as Kennedy's theory could, but it can also explain the context dependency that results of *clean* and *dirty* showed, which Kennedy's theory could not. However, the results of *full* and *empty* are not predicted by the T&S theory, as it clearly shows that absolute adjectives can also derive their standard of comparison not using a within-individual comparison class.

The theory I propose is a combination of the two theories proposed by Kennedy and T&S. I combine the two theories by specifically expanding on the T&S theory: either within-individual or between-individual comparison class is assigned to an adjective through the nature of its typical arguments, as T&S proposed, but the assignment is only a strong preference, or a tendency, rather than a permanent type that is used across all contexts. For example, absolute adjectives tend to have a within-individual comparison class due to the nature of their typical arguments, as T&S proposed, but this tendency can give way and allow the adjective to derive its standard of comparison from its scalar endpoint if context calls for it. In this way, my proposal expands on the aspect of the T&S theory that claims that all adjectives are context dependent: all adjectives are context dependent in the sense that their typical arguments determine the preferred comparison class of the adjective, as T&S proposed, but all adjectives are also context dependent in a wider sense, where the context at the utterance level can also influence whether an adjective utilizes this assigned comparison type, or utilize its scalar endpoint, if the adjective has one. This modified version of the T&S theory can account for the results we have seen in the experiment of this thesis: all relative adjectives displayed a between-individual comparison class as that is the tendency assigned to them by their typical arguments. Since relative adjectives also do not have any semantic endpoints, Kennedy's *Interpretive Economy* does not enforce an endpoint-related

interpretation with these adjectives. Two of the absolute adjectives displayed a standard of comparison derived from the endpoints in their semantic scale, but the other two absolute adjectives displayed a within-individual comparison class. This can be explained by the proposed theory if it can be said that the context in which *empty* and *full* were used allowed the adjectives to deviate from its assigned type of comparison class and use its scalar endpoint as their standard of comparison.

Not only can my take on the T&S theory explain the results of the experiment, but it also eliminates two of the concerns I proposed regarding the T&S theory by significantly decreasing the role that typical arguments play in determination of the comparison class of adjectives. My first concern regarding the T&S theory was that it was unclear how the comparison class was determined for adjectives whose typical arguments were unclear, such as *opaque*. The prediction made with the new theory would be that it simply does not have a preferred tendency, as it is not necessary, and context alone determines it every time it is used, and it can naturally have a within-individual or a between-individual comparison class. It is a closed adjective since it has a totally closed scale structure, but whether it utilizes a scalar endpoint to be the standard of comparison or adopt a standard of comparison from a comparison class ultimately depends on the context every time it is used. The second concern was what would happen if an adjective was paired with an atypical argument. This concern also disappears with this version of the T&S theory, as it would simply predict that it would adopt to the context in which the atypical argument is used. If for example the absolute adjective *clean* is used with the argument *Neptune's atmosphere*, which has an enduring property when it comes to cleanliness, then *clean* would adopt a between-individual comparison class in order to adapt to the context in which it is being used.

The results of this theory provide evidence that the derivation of the standard of comparison of gradable adjectives can be highly context dependent, in the way that the T&S theory proposed but also beyond. What we cannot conclude from the results of the experiment, however, are the details of which context factor can actually cause an absolute adjective to derive its standard of comparison from its scalar endpoint rather than its assigned comparison class. Although not conclusive, I propose that one context factor at the utterance level that could influence how the standard of comparison is derived for gradable adjectives is the context of purpose, or function.

4.6 Functional-relative approach: a possible research direction
One possible context factor that could determine whether an absolute adjective takes on its standard of comparison from its assigned within-individual class or deviate away from it to adopt one from its scalar endpoint, is the presence of functionality. Depending on the function of the subject that the adjective modifies, it can either be interpreted with respect to its given comparison class or its scale. This proposal that this type of pragmatic context influences our interpretation of gradable adjectives is not novel. Graff (2000), for instance, proposed that vague predicates are vague because our interest, or present purpose of the context, allows it to be vague: often, it is acceptable to say that different but similar things are same for the present interest or purpose. For example, imagine a situation where you are

making coffee and while transferring a scoop of coffee beans into the coffee machine, you drop a few the ground. A child who was watching you wrongly believes that the scoop is a measuring cup rather than a tool for transfer and concludes that coffee making is an exact science. When the child tells you that you must add some more beans to replace the ones you dropped, you may explain to the child that it does not matter, because the amount of coffee beans you ended up using is the same as if you had not dropped some beans for the purpose of brewing coffee. In this scenario, whether a pile of coffee beans you ended up putting in your machine is the same amount as what you initially scooped does not depend on whether it 'stands out' in quantity somehow, but whether it would fulfill the same purpose as the initial scoop. This is one example taken from Graff that shows that the purpose, or interest, of a context can guide how a predicate is used.

My proposal that function of the context can influence whether an absolute adjective takes on its standard of comparison from a comparison class or its given scalar endpoint can be seen as an application of this interest-relative theory of vagueness. Take for example the context in which *empty* was tested in this thesis: the participants had to decide whether they considered the gas tank to be empty when Emma left the gas station. This proposal would explain that the reason why almost all participants, no matter if they received story A or story B, gave the response value of 1 (disagreeing strongly that it is empty) to the statement is that when it comes to assessing the emptiness of a car's gas tank, we assess it through whether it can fulfill the function of the car. If the gas tank of a car is empty, then it will not be able to be driven. If the car can be driven, then it means that the gas tank is not empty. Since Emma drives to work after visiting the gas station, it is clear that the car moves, and therefore its gas tank is not empty. In this context where the adjective *empty* predicates over a car's gas tank, a functional-relative approach applies and therefore its standard of comparison is not derived from a within-individual comparison class but from its scalar endpoint: the function clearly draws a line that even if the gas tank is filled with less amount than it usually does, as long as it can carry out its function it is not considered empty.

However, with the case of *clean*, the functional-relative approach allows the adjective to derive its standard of comparison from its assigned within-individual comparison class. Although both stories mention that Dr. Janssen brings a surgical knife home from the hospital to use it to cut fruit, the function that the knife served the past 2 years in the two stories differ: the function of the knife in story A was always to perform surgery, while it was always to cut fruits and vegetables in story B. When the participants had to determine whether the knife was clean at "kitchen-level cleanliness", the participants may have considered the requirement to be whether it is able to perform the function that it usually performs: if a knife was previously used for surgery, then it is only ever clean at a surgery-level cleanliness. Therefore, for the adjective *clean* being predicated over a surgical knife, the functional-relative approach applies where the adjective seemingly takes on a within-individual comparison class. However, the exact interpretation of this phenomenon would be whether the knife can carry out the function that it usually does, rather than whether it purely meets the degree of cleanliness that it usually does.

The idea that a functional-relative approach determines the thresholds of adjectives in certain context can also be found in how we use for-phrases. Take (24) for instance:

(24)   This room is clean for a party.

Toledo & Sassoon claim that for-phrases co-occur with absolute adjectives only if the for-phrase references counterparts of the individual of which that adjective is predicated. While that may be the case for a lot of co-occurrences of absolute adjectives and for-phrases (Bylinina, 2013; Solt, 2009), it is not always as we can see with (24). In this case, the for-phrase *for a party* does not refer to counterparts of the room in question: we can determine whether a room is clean or not even if we have no idea how clean or dirty it usually is. What we use to determine the truth value of (24) is by considering whether the room's cleanliness meets the requirement to hold a party. In other words, we consider whether the cleanliness of the room allows it to function as a room that can hold a party.

Summarizing this discussion, the results of the current thesis showed that we cannot conclude that the comparison class is strictly within-individual for absolute adjectives and between-individual for relative adjectives, as Toledo & Sassoon proposed: if absolute adjectives were interpreted with respect to an endpoint due to a principle of grammaticalization as they proposed, rather than them taking their standard of comparison from their built-in scale structure, it cannot explain why the adjectives *full* and *empty* were unaffected by manipulation of within-individual degree presented by the story-type variable of this experiment. I proposed an alternate theory where I combined Kennedy's *Interpretive Economy* and the T&S theory as a possible explanation for the results we found in this experiment, and I also proposed a functional-relative approach in order to explain the variability of the pattern we saw in the results regarding absolute adjectives in this experiment.

5. Conclusion

This thesis investigated the semantics of the positive form of gradable adjectives, and specifically focused on the difference between Kennedy's (2007) principle of *Interpretive Economy* and the theory proposed by Toledo & Sassoon (2011) on how positive forms of gradable adjectives create a standard of comparison. Experimentally, this thesis tested the proposal made by Toledo & Sassoon that all adjectives have a semantically embedded type of comparison class, either within-individual or between-individual, depending on the nature of the property that their typical arguments hold. The experimental hypothesis was not confirmed, and therefore we could not conclude that this aspect of Toledo & Sassoon's theory is correct.

Additionally, I proposed a theory that combined Kennedy's principle of *Interpretive Economy* and the theory made by Toledo & Sassoon in order to provide a theory that could explain the results of thesis. However, it is presented as an ad hoc theory that could possibly explain the experimental results, and separate experiments must be conducted in the future if

the validity of the theory is to be confirmed. Furthermore, I proposed a functional-relative approach derived from Graff's (2000) interest-relative approach to explain the variability of the patterns we observed in the results of the experiment. This is also a theory that for its validity to be concluded, must be supported with further experimental results.

One certain finding of this thesis is that context plays a role in our interpretation of gradable adjectives, no matter whether the adjective is absolute or relative. Toledo & Sassoon had a similar approach where context of the typical arguments of gradable adjectives, no matter if the adjective is absolute or relative, assigned the type of comparison class for all adjectives. However, they proposed that ultimately the assigned type of comparison class was embedded in the semantic information of the positive form of these adjectives. The result of this thesis presents the possibility that in addition to the influence of context that Toledo and Sassoon proposed, pragmatic context at the utterance level also play a role in our interpretation of gradable adjectives as well. In this way, this thesis adds to the common theme of research in the field of semantics and pragmatics: where do we draw the line between semantics and pragmatics when it comes to the interpretation of language? Although only focusing strictly on positive forms of gradable adjectives, this thesis presents the idea that pragmatics really does matter: there seem to be other pragmatic factors, such as the function of the argument, as I have proposed in the Discussion, that ultimately influence our interpretation of language. In order to understand our interpretation of language to the most minute detail, and with the goal of understanding how we understand and interpret not just adjectives but language overall, it seems critical to consider many more pragmatic factors and see how they influence our language use in the everyday world.

6. References

Bartsch, R., & Vennemann, T. (1973). Semantic structures: A study in the relation between syntax and semantics. Frankfurt: Athäenum Verlag.

Bogusławski, A. (1975). Measures are measures: In defence of the diversity of comparatives and positives. *Linguistiche Berichte* 36: 1–9.

Bylinina, L. (2013). The Grammar of Standards: Judge-dependence, Purpose-relativity, and Comparison Classes in Degree Constructions (Doctoral dissertation, Utrecht University).

Cresswell, M. J. (1977). The semantics of degree. In B. Partee (Ed.), Montague grammar (pp. 261–292). New York: Academic Press.

Cruse, D. A. (1986). Lexical semantics. Cambridge, UK: Cambridge University Press.

Graff, D. (2000). Shifting sands: An interest-relative theory of vagueness. *Philosophical Topics,* 20: 45-81.

Kennedy, C. (2007). Vagueness and grammar: The semantics of relative and absolute gradable predicates. *Linguistics and Philosophy* 30(1): 1-45.

Leffel, T., Xiang, M., & Kennedy, C. (2016). Imprecision is pragmatic: Evidence from referential processing. *Semantics and Linguistic Theory,* 26: 836-854.

McNally, L. (2011). The relative role of property type and scale structure in explaining the behavior of gradable adjectives. In *International Workshop on Vagueness in Communication,* 151 – 168. Springer, Berlin, Heidelberg.

Paradis, C. (1997). Degree modifiers of adjectives in spoken British English. (Lund Studies in English; Vol. 92). Lund University Press.

van Rooij, R. (2011). Vagueness and linguistics. In G. Ronzitti (Ed.), *Vagueness: A Guide, Logic, Epistemology, and the Unity of Science Series,* 123-170. Springer, Dordrecht.

Rotstein, C., & Winter, Y. (2004). Total adjectives vs. partial adjectives: Scale structure and higher order modifiers. *Natural Language Semantics,* 12: 259 – 288.

Sauerland, U., & Stateva, P. (2011). Two Types of Vagueness. In *Vagueness and Language use*, 121-145. Palgrave Macmillan, London.

Solt, S. (2009). Notes on the comparison class. In *International Workshop on Vagueness in Communication*, 189-206. Springer, Berlin, Heidelberg.

Stechow, A. von (1984). Comparing semantic theories of comparison. *Journal of Semantics* 3: 1-77.

Toledo, A., & Sassoon, G.W. (2011). Absolute vs. relative adjectives – variance within vs. between individuals. *Semantics and Linguistic Theory*, 21: 135-154.

7. Appendix.

List of all testing material used in the experiment. Items number 1 to 10 are the experimental items, and items number 11 to 22 are the filler items.

Absolute adjectives

1. Clean
   a. Dr. Janssen is a medical surgeon who easily grows attached to his work tools. Today, he took a surgical knife he had been using for 2 years at the hospital for surgery back home to use it as a kitchen knife. He finds the small knife to be perfect for cutting small fruits and vegetables. After bringing it home, he washed the knife in the dishwasher along with his other kitchen knives.
   b. Dr. Janssen is a medical surgeon who easily grows attached to his work tools. 2 years ago, he took a never-used, brand new surgical knife from the hospital back home to use it as a kitchen knife. He finds the small knife to be perfect for cutting small fruits and vegetables. At home, he washes the knife in the dishwasher along with his other kitchen knives.

   Statement: After being washed in the dishwasher, Dr. Janssen's knife is clean.

2. Dirty
   a. Noah is a construction worker who has a lucky shirt he wears to work every day, believing that it keeps him safe. The first day he wore the shirt 5 months ago, he luckily avoided a fatal injury at work. Ever since, he has never washed the shirt. Yesterday, he decided to wash the shirt, but he forgot to use detergent while washing it. Although the shirt still had some spots due to the lack of detergent, Noah still decided to wear the shirt today for work. When he had got to work, he remembered about the important meeting he had to attend later today regarding his promotion.
   b. Noah is a construction worker who has a lucky shirt he wears to work every day, believing that it keeps him safe. The first day he wore the shirt 5 month ago, he luckily avoided a fatal injury at work. Ever since, he washes the shirt at the end of the day every day, so it's ready to be worn the next day. Yesterday, he forgot to use detergent while washing it. Although the shirt still had some spots due to the lack of detergent, Noah still decided to wear the shirt today for work. When he had got to work, he remembered about the important meeting he had to attend later today regarding his promotion.

   Statement: Noah wore a dirty shirt to the meeting.

3. Full

a. Olivia has a favorite mug at home which she has owned for 3 years, which she always uses when she drinks water. She always fills the mug to its brim whenever she uses it. However, after work today she came home and filled the mug up to the mid-point. At this moment, she received an emergency work call which she needed to attend to right away. She went into her room to take the call, and forgot about drinking her water.

b. Olivia has a favorite mug at home which she has owned for 3 years, which she always uses when she drinks water. She always fills the mug with only a few sips, well below half capacity whenever she uses it. However, after work today she came home and filled the mug up to mid-point before starting to drink it. At this moment, she received an emergency work call which she needed to attend to right away. She went into her room to take the call, and forgot about drinking her water.

Statement: The mug Olivia left out is full.

4. Empty
   a. Emma has been driving the same sedan for 5 years. Ever since she got the car, she has always filled the gas tank of her car with only the minimum amount of gas the tank could hold. However, last week at the gas station she filled her tank for the first time to half the tank's max capacity before going to work.
   b. Emma has been driving the same sedan for 5 years. Ever since she got the car, she has always filled the gas tank of her car with the maximum amount of gas the tank could hold. However, last week she filled her tank for the first time at half the tank's max capacity before going to work.

Statement: Emma left the gas station with an empty gas tank.

5. Wet
   a. The Regenwoud forest is a forest that has rained heavily every single day ever since it was discovered and documented 50 years ago. However, due to climate change the weather around the forest has started to change. Starting 4 years ago, it has started to rain once a week on average, while staying dry the rest of the week. Gerel have been studying the weather of the forest for his whole life. He finally had the chance to visit the forest for the first time last week.
   b. The Regenwoud forest is a forest that has never rained a single day ever since it was discovered and documented 50 years ago. However, due to climate change the weather around the forest has started to change. Starting 4 years ago, it has started to rain once a week on average, while staying dry the rest of the week. Gerel have been studying the weather of the forest for his whole life. He finally had the chance to visit the forest for the first time last week.

Statement: Gerel visited a wet forest.

6. Dry

a. The Woestijn desert is a desert that never rained ever since it was discovered and documented 50 years ago. However, due to climate change the weather around the desert has started to change. Starting 4 years ago, it has started to rain once a month on average, while staying dry the rest of the month. Gerel have been studying the weather of the desert for his whole life. He finally had the chance to visit the desert for the first time last week.

b. The Woestijn desert is a particularly moist desert which has rained on average three days a month ever since it was discovered and documented 50 years ago. However, due to climate change the weather around the desert has started to change. Starting 4 years ago, it has started to rain once a month on average, while staying dry the rest of the month. Gerel have been studying the weather of the desert for his whole life. He finally had the chance to visit the desert for the first time last week.

Statement: Gerel visited a dry desert.

Relative adjectives

7. Deep
   a. James has been working as a volunteer, digging wells in villages without water for the past year. The standard depth of a well that he was trained to dig was 7 meters. He has followed this standard procedure throughout the dozens of wells he dug the past year, except for the very first one. The first well he ever dug was 10 meters deep, which did not allow the well to be used. After a year of no usage, the well was fixed today to be the standard depth of 7 meters, in which after the well could be used for the first time.
   b. James has been working as a volunteer, digging wells in villages without water for the past year. The standard depth of a well that he was trained to dig was 7 meters. He has followed this standard procedure throughout the dozens of wells he dug the past year, except for the very first one. The first well he ever dug was 5 meters deep, which did not allow the well to be used. After a year of no usage, the well was fixed today to be the standard depth of 7 meters, in which after the well could be used for the first time.

   Statement: James' first well is deep.

8. Expensive
   a. Evan has been looking for a new couch to buy for his apartment for the past few months at his local furniture store. The couch he really wanted to buy was by far the most expensive couch in the store. Evan was saving up for 5 months for the couch. When he entered the store yesterday afternoon to check out the couch again, he saw that the couch was on sale. The couch was now the same price as most other couches in the store. Evan purchased the couch right away, before the price could increase again.

b. Evan has been looking for a new couch to buy for his apartment for the past few months at his local furniture store. The couch he really wanted to buy was very inexpensive, as there were many more expensive couches in the store. Evan was saving up for 5 months for the couch. When he entered the store yesterday afternoon to check out the couch again, he saw that the price of the couch had increased. The couch was now the same price as most other couches in the store. Evan purchased the couch right away, before the price could increase even more.

Statement: Evan's couch is expensive.

9. Long
   a. Jan is a writer who has published many novels. She had been working on her latest novel for the past 2 years. She had a version ready for publication a month ago, and it was the longest book she had written so far: it was triple the length of any other books she had published. However, her publisher wanted her to change the book so that its length matched the length of all her previous books. Jan had to comply, and her published version ended up being the same length as all the other books she had published before.
   b. Jan is a writer who has published many novels. She had been working on her latest novel for the past 2 years. She had a version ready for publication a month ago, and it was the shortest book she had written so far: it was half the length of any other books she had published. However, her publisher wanted her to change the book so that its length matched the length of all her previous books. Jan had to comply, and her published version ended up being the same length as all the other books she had published before.

Statement: Jan's latest novel is long.

10. Tall
   a. The clock tower standing in the center of the city of Dorp is as tall as any other clock towers in the country. The tower used to be the tallest clock tower in the country for 200 years, but due to a fire 5 years ago it was burned down and destroyed. The clock tower was rebuilt since then, to a height which equals all other clock towers in the country.
   b. The clock tower standing in the center of the city of Dorp is as tall as any other clock towers in the country. The tower used to be the shortest clock towers in the country for 200 years, but due to a fire 5 years ago it was burned down and destroyed. The clock tower was rebuilt since then, to a height which equals all other clock towers in the country.

Statement: The clock tower in the city of Dorp is tall.

Fillers with expected medium values

11. Emica goes out with her friends quite often to local bars. The bar that they frequent the most often went under some management changes the past few months. Although nothing much has changed, the type of music that they play at the bar became very different. Emica and her friends started to argue amongst themselves if they should continue to visit this bar, or look for a different one that could suit them better.

    Statement: Emica and her friends should find a different bar to visit regularly.

12. Roland is a paleontologist who had gotten used to desk work. He had done some field work during his studies and during his first job as a paleontologist, but for the past 6 years he had a job at a university as a teacher. However, he was notified yesterday that he would be taking on a new job, as he was being sent to do some field work for the next 6 months for the university. Roland was nervous. He was in a way excited to get out of the office but was doubting if he was still qualified to go back to doing fieldwork.

    Statement: Roland is happy to have received his new job.

13. Tanya and her twin sister were always fighting when they were teenagers. Although they had some big fights when they were younger, as they grew up things started to change. They attended different universities, and they did not keep in touch frequently. Now, they live in different countries and do not see each other often. However, their relationship is much better than when they were younger. Even though they do not see each other often still, they can be themselves and get along very well when they do.

    Statement: Tanya and her sister are close.

14. Vivek had been wanting her driving license for a long time, but her mother was too scared to let her get her license. Vivek has always been a clumsy child, and 3 years ago she was involved in a car crash which made her mother scared for Vivek's life. Throughout her high school years, Vivek was never allowed to get her license. This made Vivek's social life difficult, as she always had to depend on her mother every time she wanted to go anywhere.

    Statement: Vivek's mother was right to worry about Vivek as she did.

15. Marc has always loved to read. Ever since he was young, he would visit the local library at least once a week and read through several books every week. However, he rarely buys books. He always rents them from the library. This is nice for Marc because he does not make much money as a teenager. However, the library is quite far from the house and he has to walk there every week. He has some money saved up that he could use to buy some books. But Marc doesn't feel the need to buy them if he can borrow them from the library.

    Statement: Marc should buy some books of his own.

16. Frank picked up his daughter from kindergarten. Frank needed to take her to the dentist, and to convince her Frank told her that he would take her to the mall and buy

her ice cream after if she came along nicely. His daughter did so, but as they were at the dentist it started to storm outside. Although Frank had promised her ice cream, he did not want to drive out in the storm. Frank did not keep his promise and took his daughter straight home after she was done. His daughter was sad and cried, but she understood why Frank did not take her to the mall.

Statement: Frank should not have promised his daughter ice cream.

Fillers with expected high values

17. Elise is a plumber who has been working in the city for the past 10 years. However, she had always wanted to become a musician, and dreaded working as a plumber. Throughout the past 15 years, she dedicated all her free time into learning how to play the piano, one day hoping to make a living as a jazz pianist. One day, she attended an audition to be part of a jazz group and was selected. She could now become a full-time musician if she wanted to, but that meant she had to stop working as a plumber.

    Statement: Elise wants to become a jazz pianist.

18. Jimmy has been in the hospital for the past few months, as he was involved in a major car crash. He had broken both of his legs and had a serious concussion. To make matters worse, his car was completely destroyed during the accident. However, his accident was all over the news and his unfortunate story was known by all of his friends and family. They had secretly bought Jimmy a new car, and were waiting to give the car to him once Jimmy was fully recovered.

    Statement: Jimmy broke both of his legs due to the accident.

19. Louis went to his favorite coffee place. For the past 4 years, he always had a black coffee to go on his way to work. This morning he did the same, but received a tea instead of his usual coffee. He was unhappy about the mistake, but he was in a hurry, so he did not complain and took the tea. He was pleasantly surprised when he tasted the tea, as it was much better than what he was expecting. He thought about getting the tea more often instead of his regular coffee.

    Statement: Louis enjoyed his tea this morning.

Fillers with expected low values

20. Eve has been living by herself ever since she became an adult. This was a big transition for her as she is legally blind. After 5 years of living alone, she had mastered her way around her house. She did not need any help in carrying out her

daily routine. However, she was starting to feel alone living by herself, so she adopted a dog to keep her company.

Statement: Eve needed a seeing-eye dog to help with her daily routine.

21. Ronnie got into a huge argument with her boyfriend who she has been dating for the past 2 years. Although she liked him a lot in the beginning, Ronnie had felt that the relationship was falling apart starting half a year ago. When they were having dinner at a restaurant last night, her boyfriend was being very rude to their waiter. That behavior from him was the last straw for Ronnie, and she did not want to see him anymore. He called her when she got home to talk things through, but Ronnie didn't know if the relationship was worth saving.

Statement: Ronnie was never into her boyfriend that much.

22. Itsuki lives with 2 housemates. In their apartment, they had a flatscreen television that Itsuki's parents gave to her as a gift when she moved out to live independently. The television was used for 4 years before it broke down yesterday. Itsuki and her housemates talked about buying a new television, and her housemates wanted to divide the cost equally among the three of them. However, Itsuki was planning to move out in 2 months, and did not want to split the cost equally. She did not know how to tell her housemates that she was planning to move out so soon.

Statement: Itsuki wants to split the cost of the new television equally with her housemates.