CHAPTER 1

Meaning Tracks Use

Meaning tracks use. If an overwhelming majority of competent language users frequently *say* that some acts are somewhat right and somewhat wrong, then this indicates that RIGHT and WRONG *are* gradable concepts. Obviously, for this argument to be convincing, it is not enough to show that a few individuals occasionally use RIGHT and WRONG as gradable concepts. Not every usage of a concept is thoughtful and sincere. What we need to show is that many people *persistently* and *sincerely* talk about RIGHT and WRONG as if they permit of degrees.[1] The structure of the meaning-tracks-use argument is, thus, as follows: (1) If the vast majority of competent language users frequently and sincerely *use* RIGHT and WRONG as gradable concepts, then RIGHT and WRONG *are* gradable concepts. (2) The antecedent of the first premise is true. (3) Therefore, RIGHT and WRONG are gradable concepts.

On a strict interpretation of the first premise, the vast majority of competent language users cannot be wrong about their persistent and sincere use of moral concepts. This is, however, not the intended interpretation. The meaning-tracks-use argument merely assumes that the way people use words and phrases is a *reliable guide* to meaning. This assumption is weaker than Wittgenstein's claim that meaning *is* use.[2] Regardless of what meaning is, few would question that we can typically figure out the meaning of a concept by studying how it is used. This holds true for moral as well as nonmoral concepts.

The epistemic interpretation of the meaning-tracks-use argument is endorsed by modern linguists. For instance, Katrin Erk notes that so-called vector space models of meaning are based on the observation that

[1] As noted on p. 4, the gradualist hypothesis has a semantic as well as a non-semantic part. This chapter is exclusively concerned with the semantic component of the gradualist hypothesis.

[2] See Wittgenstein (1953) and the quote in footnote 30 on p. 11.

"we can often guess what a word means from the contexts in which it is used. Thus, we can represent meaning as distribution, as observed contexts."[3] Vector space models have proven to be empirically successful. By observing how words are used in large corpora, by counting their occurrence in different sentences, computational linguists have developed software that accurately predicts whether the meaning of two words is similar or not.[4] However, a well-known limitation of the computational approach is that it requires large quantities of empirical data. As Erk puts it, "many phrases do not occur with sufficient frequency in a corpus to be represented through their distributional contexts."[5] This includes phrases relevant for assessing the gradualist hypothesis, for example, "this act is somewhat right and somewhat wrong" and "this act is a bit right and a bit wrong." To overcome this problem, this chapter uses the methods of experimental philosophy. Data from three surveys are presented. In the largest, which had more than 700 respondents, no more than four percent of ordinary language users persistently used RIGHT and WRONG as binary concepts.[6] The statistical analysis also indicates that RIGHT and WRONG are used as gradable concepts to approximately the same extent as color concepts, which suggests that rightness and wrongness come in degrees about as much as colors do. Furthermore, by using multidimensional scaling techniques, it can be shown that RIGHT and WRONG are used as gradable concepts even when no gradable terms (such as "right to some degree" or "somewhat right and somewhat wrong") appear in the questions or answer options. This reduces the risk of acquiescence bias.

The presentation of the empirical evidence for the meaning-tracks-use argument presupposes some knowledge of statistics. Readers who are willing to accept the claim that the vast majority of competent speakers do use RIGHT and WRONG in ways that permit of degrees without scrutinizing the empirical evidence can skip the rest of this chapter.

[3] Erk (2012: 635).

[4] A popular measure of similarity in linguistic is the cosine of the vector space representation of each word-pair. For details, see Erk (2012: 637) and Turney and Pantel (2010: 160).

[5] Erk (2012: 637). According to Google's word2vec algorithm the five most similar words to "moral" are "ethical" (cosine distance 0.363), "ethics" (0.471), "religious" (0.489), "morality" (0.513), and "philosophical" (0.518). It is not surprising that "ethical" is more similar to "moral" than "religious," but some readers might find it surprising that "religious" is closer to "moral" than is "philosophical."

[6] This figure is a summary of the overall results based on data from all semiabstract and concrete items included in the largest of the three studies, which included four answer options. The corresponding figure for the second study, in which each respondent evaluated only one semiabstract item and was presented with five answer options was, as expected, higher (twelve percent).

Background and Design

Questionnaires were distributed to three sets of respondents: students at Texas A&M University taking a class in engineering ethics in Spring 2019 (n = 715), students taking the same class in Spring 2020 (n = 578), and a group of U.S. citizens who voted in the 2016 election (n = 182). Students took the surveys for credit (about 0.5% of the total course grade) while respondents in the third study received between $2.65 and $3.00. In all three studies, respondents were invited to answer up to twelve questions. The order of the questions as well as the answer options was randomized. At the request of the Institutional Review Board at Texas A&M University, no demographic information was collected.

Respondents were presented with up to five different types of tasks: abstract, semiabstract, concrete, comparative, and open-ended tasks. All tasks are described in the following sections. By using numerous measures for testing a single hypothesis, the validity of the measurement instrument can be assessed. If the validity is high, we should expect the results to be roughly the same for all types of tasks.

Abstract Tasks

Subjects were invited to evaluate abstract statements about ethics, mathematics, colors, scientific evidence, and scientific facts on a seven-point Likert scale ranging from "strongly disagree" (1) to "strongly agree" (7). The following example pertains directly to the gradualist hypothesis:

A1. Moral rightness and wrongness come in degrees. The boundary that separates morally right acts from wrong ones is not always sharp. Some acts are somewhat right and somewhat wrong.

Study 1: average degree of agreement 5.7 (n = 237, std. dev. 1.4)
Study 2: average degree of agreement 5.4 (n = 114, std. dev. 1.3)
Study 3: average degree of agreement 5.4 (n = 181, std. dev. 1.5)

The following abstract statements were used as reference points for comparative purposes:

A2. In mathematics, truth comes in degrees. The boundary that separates true mathematical statements from false ones is not always sharp. Some mathematical statements are somewhat true and somewhat false.

Study 1: average degree of agreement 2.7 (n = 219, std. dev. 1.7)
Study 2: average degree of agreement 2.6 (n = 79, std. dev. 1.6)
Study 3: average degree of agreement 2.2 (n = 181, std. dev. 1.5)

https://doi.org/10.1017/9781009336772.003 Published online by Cambridge University Press

Table 1.1 *Mann–Whitney* U*-tests for data in Study 2*

	A2. In mathematics, truth comes in degrees … (n = 79)	A3. Colors come in degrees … (n = 80)
A1. Moral rightness and wrongness come in degrees … (n = 114)	Mann–Whitney U = 1,111 p-value < 0.00001. Significant at p < 0.01; one-tailed.	Mann–Whitney U = 4,014 p-value = 0.0778. Not significant at p < 0.05; one-tailed.

A3. Colors come in degrees. The boundary that separates one color from another is not always sharp. Some color hues are somewhat red and somewhat blue.

Study 1: average degree of agreement 5.9 (n = 233, std. dev. 1.1)
Study 2: average degree of agreement 5.7 (n = 80, std. dev. 1.2)
Study 3: average degree of agreement 6.2 (n = 182, std. dev. 1.2)

All items were followed by a simple comprehension check. Responses from subjects who did not answer it correctly were excluded from the analysis.

The data sets are not normally distributed. It is therefore appropriate to perform a Mann–Whitney *U*-test. This is a nonparametric test for the null hypothesis that the distributions of two data sets are identical and therefore have the same median value. Table 1.1 summarizes the results for Study 2, which is less likely than the others to yield significant results due to its smaller sample size. The Mann–Whitney *U*-test indicates that moral rightness and wrongness is judged to come in degrees to a significantly higher extent ($p < 0.01$, one-tailed) than truth in mathematics is judged to come in degrees, but there is no statistically significant difference between A1 (moral rightness and wrongness come in degrees) and A3 (colors come in degrees), not even at $p < 0.05$. This indicates that rightness and wrongness is judged to come in degrees to approximately the same extent as colors are judged to come in degrees. The results of Study 1 offer additional support to this conclusion: all comparisons in Study 1 yield significant results ($p < 0.01$, one-tailed) except that between colors and moral rightness. However, in Study 3 the comparison between colors and moral rightness yields a significant difference ($p < 0.01$, one-tailed), indicating that the extent to which colors and moral rightness come in degrees is not *exactly* the same.

The abstract tasks also included the following general statements about scientific evidence and scientific facts:

https://doi.org/10.1017/9781009336772.003 Published online by Cambridge University Press

Table 1.2 *Four ordinal levels of agreement in Study 1 and Study 2. All pair-wise differences between items at different levels are statistically significant at* p < *0.01; one-tailed*

Level 1	Colors come in degrees …	Moral rightness and wrongness come in degrees …
Level 2	Scientific evidence comes in degrees …	
Level 3	In science, facts come in degrees …	
Level 4	In mathematics, truth comes in degrees …	

A4. Scientific evidence comes in degrees. The boundary that separates theories corroborated by evidence from those that are not is not always sharp. Some scientific theories are somewhat supported by evidence and somewhat unsupported.

Study 1: average degree of agreement 4.2 (n = 239, std. dev. 1.8)
Study 2: average degree of agreement 4.8 (n = 97, std. dev. 1.4)
Study 3: average degree of agreement 4.3 (n = 182, std. dev. 1.8)

A5. In science, facts come in degrees. The boundary that separates correct scientific claims from incorrect ones is not always sharp. Some scientific claims are somewhat correct and somewhat incorrect.

Study 1: average degree of agreement 3.4 (n = 236, std. dev. 1.8)
Study 2: average degree of agreement 4.1 (n = 68, std. dev. 1.7)
Study 3: average degree of agreement 3.9 (n = 185, std. dev. 1.8)

In all three studies, the Mann–Whitney U-test for A4 vs. A5 indicates that scientific evidence is judged to come in degrees to a somewhat higher extent than scientific facts (in Study 2, U = 2,615.5, p-value < 0.00001; significant at p < 0.01; one-tailed). Scientific facts are also judged to come in degrees to a significantly higher extent than mathematical truths in all three studies (in Study 2, U = 1,454.5, p-value < 0.00001; significant at p < 0.01; one-tailed), and moral rightness and wrongness are reported to come in degrees to a significantly higher extent than scientific evidence in all three studies (in Study 2, U = 3,792.5, p-value = 0.00004; significant at p < 0.01; one-tailed). Finally, colors are reported to come in degrees to a significantly higher extent than scientific evidence in all three studies. (In Study 2, U = 2,317, p-value < 0.00001; significant at p < 0.01; one-tailed.)

These findings can be represented in a hierarchical order with four ordinal levels. Table 1.2 summarizes the results for Study 1 and Study 2. (As noted, in Study 3 colors come in degrees to a slightly higher extent

Table 1.3 *Disagreement does not explain gradualist responses*

	N	Avg.	Std. dev.
It is common that people disagree on moral issues.	114	6.1	1.0
Moral rightness and wrongness sometimes come in degrees even when there is no disagreement.	114	5.2	1.3
It is common that people disagree on mathematical issues.	79	3.2	1.6
Truth in mathematics sometimes comes in degrees even when there is no disagreement.	79	3.2	1.6

than moral rightness and should therefore be represented on a separate sublevel.) The difference in agreement in Table 1.2 between all pairs of levels is statistically significant at $p < 0.01$, except that between scientific evidence and scientific facts which is significant at $p < 0.05$.

An alternative explanation of the results in Table 1.2 could be that respondents tend to merely report their belief in how much *disagreement* there is in a certain domain, not that the phenomena themselves come in degrees. To control for this, the following items were included in Study 2 (Table 1.3).

These numbers indicate that respondents do believe that there is more disagreement on moral issues than on mathematical issues ($U = 768$, p-value < 0.00001; significant at $p < 0.01$; one-tailed). This is not surprising. However, respondents *also* believe that moral rightness and wrongness come in degrees even when there is no disagreement, and they do so to a much higher extent than for mathematical issues ($U = 1{,}587.5$, p-value < 0.00001; significant at $p < 0.01$; one-tailed). This casts doubt on the alternative explanation, but fits well with the gradualist hypothesis. There is little reason to think that respondents merely reported their belief in how much disagreement there is in a given domain.

Semiabstract Tasks

All three studies included two semiabstract tasks in which subjects were invited to complete moral statements by selecting one of a set of predefined alternatives. The first of these tasks ($n = 242$, 287, and 90) was formulated as follows:

https://doi.org/10.1017/9781009336772.003 Published online by Cambridge University Press

S1. Lying in a situation in which doing so would bring about the best consequences is …

	Study 1	Study 2	Study 3
… always morally right.	2.8%	12.6%	2.2%
… always morally wrong.	6.6%	0.7%	10.0%
… sometimes right to some degree, but also wrong to some degree.	89.3%	75.0%	58.9%
… always either right or wrong but sometimes more importantly right or more seriously wrong.	NA	5.8%	NA
… either right or wrong, there is no middle position, but the truth of a moral judgment is relative to one's moral standard or cultural norms.	NA	NA	17.8%
… sometimes right and sometimes wrong, but never a bit right and a bit wrong.	NA	NA	6.7%
… not possible to assess.	1.2%	5.8%	NA
… I don't know if this is right or wrong.	NA	NA	4.4%

The gradualist option was the most frequently selected answer option in all three studies. The differences in gradualist responses in Study 1 (88.2%), Study 2 (75.0%), and Study 3 (58.9%) is explained by the fact that respondents were presented with four answer options in Study 1, five in Study 2, and six in Study 3. In Study 3 about 17.8% chose the relativist answer option, which was not available in the other studies. The more options respondents are offered to choose from, the less likely is it that everyone selects the same option. If RIGHT and WRONG had been binary concepts, many subjects could have been expected to favor the Kantian answer option ("… always morally wrong") or the utilitarian answer option ("… always morally right").

Study 2 included an answer option designed to capture Thomas Hurka's notion of degrees mentioned in the introductory chapter, according to which some acts are more importantly right or more seriously wrong.[7] This option was selected by no more than 5.8% of respondents.

In Study 1, a separate group of respondents was presented with the answer option "… sometimes a bit right and a bit wrong" instead of "… sometimes right to some degree, but also wrong to some degree." About 83.6% (n = 116) selected "… sometimes a bit right and a bit wrong," compared to 89.3% (n = 244) for the group presented with the option "… sometimes right

[7] Hurka (2019). See also the discussion at the end of Chapter 2.

https://doi.org/10.1017/9781009336772.003 Published online by Cambridge University Press

to some degree, but also wrong to some degree." The difference between 89.3% and 83.6% is not significant ($X^2 = 8.396$, p is < 0.039; not significant at $p < 0.01$). This is an indication of robustness. The results reported here do not depend on minor alterations of the wording of the gradualist hypothesis. This finding also suggests that A BIT RIGHT AND A BIT WRONG has the same meaning as RIGHT TO SOME DEGREE, BUT ALSO WRONG TO SOME DEGREE, although more data would be needed before a definitive conclusion could be drawn.

The vast majority of respondents also reported gradualist responses for the following semiabstract item, S2 (n = 119, 288, and 90):

S2. Exceeding the speed limit in an emergency is …

	Study 1	Study 2	Study 3
… always morally right.	5.9%	4.5%	15.6%
… always morally wrong.	1.7%	7.9%	4.4%
… sometimes right to some degree, but also wrong to some degree.	88.2%	75.3%	48.9%
… always either right or wrong but sometimes more importantly right or more seriously wrong	NA	6.9%	NA
… either right or wrong, there is no middle position, but the truth of a moral judgment is relative to one's moral standard or cultural norms.	NA	NA	6.7%
… sometimes right and sometimes wrong, but never a bit right and a bit wrong.	NA	NA	23.3%
… not possible to assess.	4.2%	5.5%	NA
… I don't know if this is right or wrong.	NA	NA	1.1%

A possible explanation of why so many respondents in Study 3 selected the binary response "sometimes right and sometimes wrong, but never a bit right and a bit wrong" (23.3%) or "always morally right" (15.6%) might be that speeding, unlike lying, is viewed as less morally problematic by experienced drivers. Respondents in Study 3, U.S. citizens who voted in the 2016 election, are on average older than college students and thus more likely to drive.

Concrete Tasks

Study 1 included six concrete tasks in which respondents were invited to assess brief descriptions of particular acts. These tasks were not designed to

https://doi.org/10.1017/9781009336772.003 Published online by Cambridge University Press

study respondents' views on moral relativism or Hurka's hypothesis, so the number of answer options was limited to four. In Study 1, the following task was evaluated by all respondents (n = 715):

C1. John lies to a future employer about his qualifications. Due to inadequate background checks, his lies go undetected. He is offered, and accepts, a job he is not qualified for.

What John did was morally right.	1.1%
What John did was morally wrong.	91.3%
What John did was right to some degree, but also wrong to some degree.	6.0%
What John did cannot be assessed from a moral point of view.	1.6%

Although relatively little information is provided in the vignette, over ninety percent reported that John's act was wrong. The next item (C2, n = 363) serves as a reference point for what seems to be a case of someone doing something right:

C2. Jared's colleague Bob struggles to understand a new task for work. Jared has no plans for the evening and volunteers to help Bob to get up to speed. Between 5 pm and 8 pm, Jared helps Bob to figure out how to solve the new task.

What Jared did was morally right.	74.9%
What Jared did was morally wrong.	1.9%
What Jared did was right to some degree, but also wrong to some degree.	2.5%
What Jared did cannot be assessed from a moral point of view.	20.6%

The gradualist hypothesis states that *some* acts are somewhat right and somewhat wrong, not that *all* are. Therefore, the findings for C1 and C2 neither refute nor confirm the gradualist hypothesis. However, data for the following items, C3–C6, support the gradualist hypothesis. For these items, the gradualist answer option "What [the agent] did was right to some degree, but also wrong to some degree" was the most frequently selected answer in all three studies. See Table 1.4.

C3. Denise is in severe pain. She asks her spouse Adam to drive her to the hospital for treatment. Although he knows that her condition is not life-threatening, Adam drives 15 miles above the speed limit to bring Denise to the hospital as fast as he can.

C4. An experienced airline captain flies through a volcanic ash cloud that causes the engines to malfunction. To prevent panic among the passengers, the captain decides to lie to the passengers: "The airport at our destination is closed due to bad weather. We will land at a nearby airport. There is no danger."

Table 1.4 *Relative frequencies for items C3–C6. The answer options were the same as for C1 and C2*

	C3 (n = 329)	C4 (n = 368)	C5 (n = 353)	C6 (n = 386)
Right	10.9%	14.4%	5.7%	3.6%
Wrong	13.6%	15.5%	23.5%	37.0%
Degree	69.3%	71.5%	62.9%	43.0%
Not Assessable	6.1%	1.4%	8.0%	12.9%

C5. Anna is proud of a new webpage she has designed for a client, and the client is also pleased with it. Anna asks her junior colleague Josh to share his opinion. Josh does not like the webpage, but because Anna and the client seem to like it Josh decides to lie: "I think your new webpage looks great, congratulations!"

C6. Miriam is on her way to the airport. Due to severe congestion on the highway, she realizes that there is a risk she will miss her international flight. When the road finally clears, she drives 15 miles above the speed limit to make up for the time lost because of the congestion.

Table 1.5 summarizes pair-wise chi-square tests for all combinations of C1–C6. The degrees of freedom for all comparisons are df = 3, so for $p < 0.01$, the critical chi-square value is 11.34, and for $p < 0.001$, it is 16.27. Note that *all* pair-wise comparisons are statistically significant. However, the chi-square values for C1 and C2 stand out: they are ten to one hundred times higher than the values for all other items. (See the dashed box in Table 1.5.) From a statistical point of view, the explanation is that C3–C6 are items in which the gradualist answer option is the most popular one; therefore, C3–C6 have more in common with each other than with C1 and C2.

It is also worth noting that the chi-square values for C6 are four to ten times higher than the corresponding values for C3, C4, and C5. What could explain this? The best explanation seems to be that in items C3, C4, and C5 a widely accepted norm is violated for a good reason: By violating the speed limit, or by lying, the agent brings about good consequences for others. A large majority reported that such norm violations are somewhat right and somewhat wrong. However, in C6, the agent violates a norm for what appears to be a selfish reason. Fewer subjects considered this to be somewhat right and somewhat wrong, and about twice as many considered it to be wrong.

If we combine the findings for the abstract, semiabstract, and concrete tasks we find that only four percent in the largest study (Study 1, n = 715) persistently used RIGHT and WRONG as non-gradable concepts.

https://doi.org/10.1017/9781009336772.003 Published online by Cambridge University Press

Table 1.5 *Pairwise chi-square tests for all combinations of items C1–C6. For* p < *0.01 the critical chi-square value is 11.34, and for* p < *0.001 it is 16.27 (df = 3). Note that* all *comparisons are statistically significant at* p < *0.001*

	C1	C2	C3	C4	C5
C2	18,840.98				
C3	3,645.44	1,804.09			
C4	3,168.97	2,235.24	56.29		
C5	1,830.94	3,503.68	33.32	82.6	
C6	658.66	5,462.24	218.55	330.83	105.72

Comparative Tasks

Study 1 included a fourth type of task designed to test the gradualist hypothesis *without* including gradualist terms such as "degree" and "somewhat right" in the vignettes. If gradualist terms appear in the questions or answer options, respondents might be more willing to apply such terms than they otherwise would. Psychologists call this phenomenon acquiescence bias.[8]

Subjects were asked to make pair-wise comparisons between items C1 and C6 in Section 6 and a seventh item C7:

C7. After graduation, Zofia decides to do unpaid volunteer work for Engineers without Borders for a couple of months before joining Petersen Consulting in Dallas, TX.

The comparative task was formulated as follows:

Assess the acts performed by the agents. How similar are the moral properties of the two acts? (If one is right and the other is wrong, they are not very similar, but if both are right, or both are wrong, they are very similar.)

[C 1]

[C 2]

[Seven-point Likert scale, ranging from "very similar" to "very dissimilar."]

Pair-wise comparisons of seven items, C1–C7, require twenty-one comparisons. Each respondent was invited to make four comparisons, which yielded 2,830 comparative data points (n = 98–141). To verify that subjects understood the comparative task correctly, three identical comparisons

[8] See Messick and Jackson (1961). See also the discussion at the end of this chapter.

https://doi.org/10.1017/9781009336772.003 Published online by Cambridge University Press

Table 1.6 *Average degree of dissimilarity, ranging from 0 (very similar) to 6 (very dissimilar)*

	C1	C2	C3	C4	C5	C6	C7
C1	0.1						
C2	5.5	N/A					
C3	3.6	3.8	N/A				
C4	3.6	4.1	3.0	N/A			
C5	3.4	4.3	3.0	2.6	N/A		
C6	3.7	5.0	2.2	3.7	3.2	0.2	
C7	5.7	1.5	4.2	4.6	4.5	5.2	0.3

were included in the questionnaire, C1–C1, C6–C6, and C7–C7. The average dissimilarities reported for these items were 0.1, 0.2, and 0.3, which indicates a good understanding of the task. Table 1.6 summarizes the results.

Dissimilarities can be interpreted as distances in an *n*-dimensional geometric space. The more dissimilar two items are, the farther apart is their location. The twenty-one comparisons listed in Table 1.6 can range over twenty dimensions, but by applying multidimensional scaling techniques, the dimensionality of this multidimensional data set can be significantly reduced.[9]

The aim of a classic multidimensional scaling is to represent the original data set by a new set of points in a smaller number of dimensions such that the Euclidean distance between each pair of points in the new set approximates the distance in the original multidimensional data set. Ideally, each pair-wise distance (similarity) in the original data set (Table 1.6) should be exactly the same as the corresponding Euclidean distance in the new representation. However, as we reduce the number of dimensions, some minor errors will typically be introduced into the new representation. This is acceptable as long as the errors are small.

Figure 1.1 shows a classic multidimensional scaling of Table 1.6. The maximum error is 0.36 units, which is a relatively large error. (This worry is addressed by the next figure.) When interpreting Figure 1.1, it is important to keep in mind that item C1 is almost unanimously (91.3%) considered to be an example of wrongdoing, whereas C2 is widely considered (74.9%) to be an example of an agent doing something right. It is thus not surprising

[9] See Kruskal and Wish (1978).

https://doi.org/10.1017/9781009336772.003 Published online by Cambridge University Press

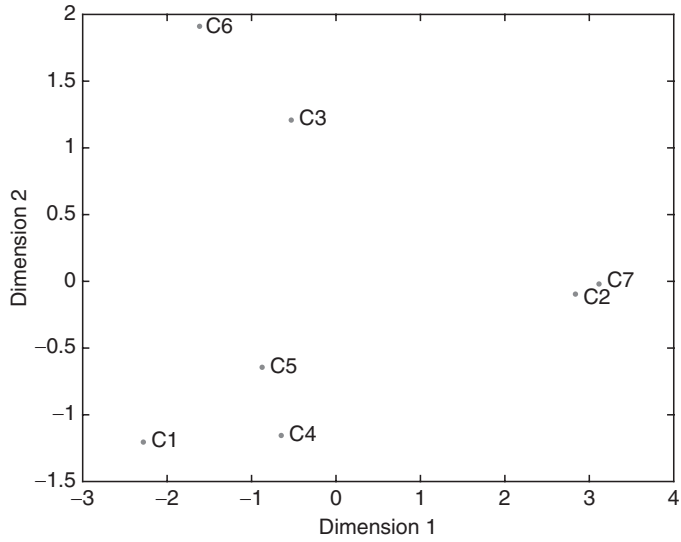

Figure 1.1 A classic multidimensional scaling of Table 1.6

that C1 and C2 are located far apart along the *x*-axis. C7 can also be taken to be an example of someone doing something right, so it is equally unsurprising that C7 is located close to C2. The locations of C1, C2, and C7 can thus be explained *without* invoking the hypothesis that rightness and wrongness come in degrees. However, the locations of C3–C6 cannot be easily explained without assuming that rightness and wrongness come in degrees. Although these items are located *somewhat* to the left in Figure 1.1, they are not close to C1 along the *x*-axis. Moreover, recall from the discussion of the concrete items that C3–C6 are items in which it is widely believed that the agent acted somewhat right and somewhat wrong (69.3% for C3, 71.5% for C4, 62.9% for C5, and 43.0% for C6). This fits well with their locations in Figure 1.1: items C1–C6 are, literally speaking, located "between" the entirely wrong act in item C1 and the entirely right acts in C2 and C7.

If the binary theory had been true, it would have been possible to represent all seven items along a single dimension, and all items would have been clustered in two distinct areas in the figure: RIGHT and WRONG. However, the findings in Table 1.6 cannot be represented in such a way and are thus incompatible with the binary theory.

That said, it remains to explain why C6 and C3 in Figure 1.1 are located above C4 and C5 on the *y*-axis. Note that C3 and C6 are cases in which

https://doi.org/10.1017/9781009336772.003 Published online by Cambridge University Press

the agent violates the speed limit for a good reason, whereas C4 and C5 are cases in which the agent lies for what seems to be a good reason. This suggests the following somewhat speculative interpretation: The more similar the agent's *reasons* for some acts are, the closer are their locations on the *y*-axis. If so, the underlying similarities between C3 and C6, and C4 and C5, seem to be visible in the figure, which indicate that data Table 1.6 have a reasonable degree of validity.

A drawback of the two-dimensional scaling is that the maximum error in Figure 1.1 is, as noted, relatively large. It is therefore appropriate to consider the three-dimensional scaling depicted in Figure 1.2. In this representation, Kruskal's stress value is 1.07×10^{-6}, which indicates a good fit. Figure 1.2 confirms the conclusion of Figure 1.1: C1 (the entirely wrong act) is located far apart from C2 and C7 (the entirely right acts), but C3–C6 are located *between* the entirely right and entirely wrong items. In Figure 1.2, it is thus also reasonable to interpret the *x*-axis as a visual representation of an act's degree of rightness. The interpretation of the *y*-axis is the same as in Figure 1.1, but I will leave it open how the *z*-axis is to be interpreted as that is of no importance to the present discussion.

Figures 1.1 and 1.2 are based on the assumption that all similarities in Table 1.6 can be represented in a metric space. Because it is hard to know if this assumption is true, nonmetric multidimensional scaling techniques are also worth considering. In this type of representation, distances are interpreted as ordinal orderings: the aim is to preserve ordinal information about the original distances in a lower number of dimensions. Figure 1.3 shows a two-dimensional nonmetric multidimensional scaling of Table 1.6. Kruskal's stress value is 8.5×10^{-5}, which indicates a good fit. This figure confirms the previous conclusions: C3–C6 is located *between* C1 and C2 & C7, which tallies well with the conclusion that the act in C1 is entirely wrong, while the acts in C3–C6 are somewhat right and somewhat wrong, and the acts in C2 and C7 are entirely right.

In summary, all three multidimensional representations fit well with the gradualist hypothesis, but they are incompatible with the binary theory.

A Socratic Midwifery Effect?

Study 3 included an open-ended task in which half of the respondents were invited to write a couple of sentences about a case without relying on any predefined answer options. The other half was invited to select one of the following answer options: "right," "wrong," "right to some degree, but also wrong to some degree," and "I don't know, I would need

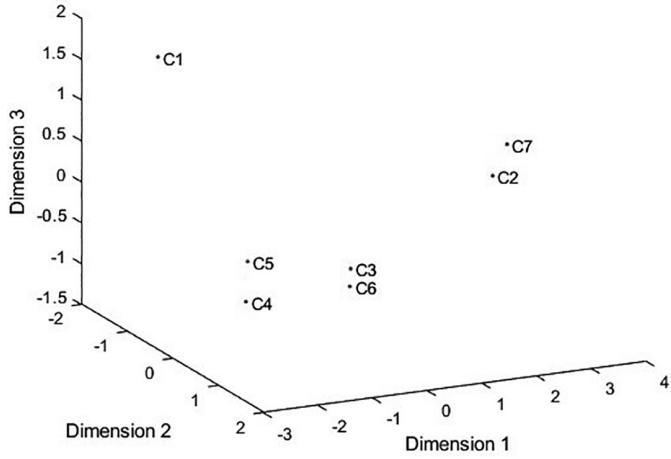

Figure 1.2 A three-dimensional metric multidimensional scaling of Table 1.6

more information." The purpose was to study to what extent gradualist answers occur naturally. The study was conducted in May 2021 during the COVID-19 pandemic:

> Anouska's 80-year-old father is reluctant to take the COVID vaccine because he is concerned about possible side effects. She explains to him that it is primarily younger women who have been affected by severe side effects, so for him the benefit would definitely exceed the risk. In Anouska's opinion, it would be irrational of her father to not take the vaccine. He eventually gives in and schedules an appointment, but only after Anouska pressures him to do so. She also deceives him by exaggerating the benefits and minimizes some of less severe side effects. Evaluate Anouska's behavior from a moral point view. What is your moral conclusion all things considered?

The first group of respondents was invited to answer the question by writing a couple of sentences without relying on any predefined answer options. That group (n = 87) submitted 3,453 words, which were analyzed and categorized manually. Gradualist conclusions were expressed by 16.1% of respondents, compared to 46.5% (n = 95) in the group presented with predefined answer options. (In the first group, 29.0% reported that it was right to pressure Anouska's father to take the vaccine, compared to 16.1% in the second group; 26.0% concluded that it was wrong to pressure Anouska's father to take the vaccine, compared to 33.3% in the second group; and 29.0% presented with the open-ended task stated views that were ambiguous or could not be reasonably classified as all-things-considered verdicts,

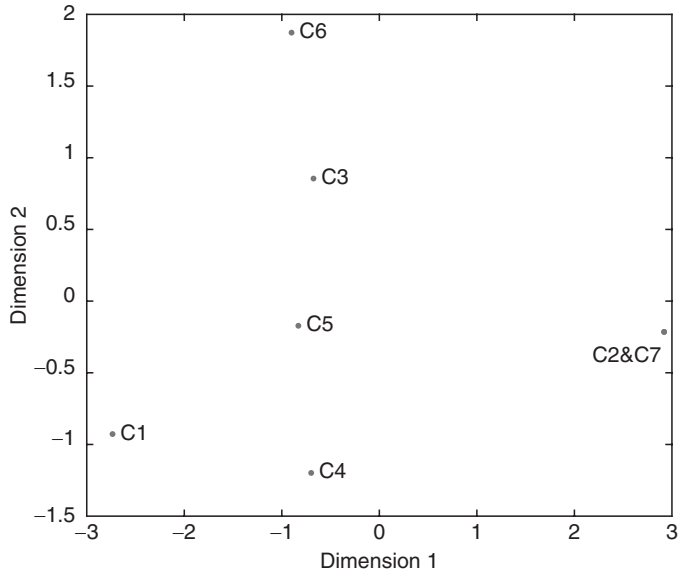

Figure 1.3 A two-dimensional nonmetric multidimensional scaling of Table 1.6

compared to 4% in the second group who selected "I don't know, I would need more information.") Here are some examples of spontaneously submitted gradualist answers from respondents in the first group:

> "What she did was morally grey, but the end result is positive."
> "From a moral standpoint, it would probably be both right and wrong, but erring toward wrong."
> "It's a little right and a little wrong."
> "It is a bit right and wrong to do what she did."

It is not surprising that respondents used gradualist terms spontaneously, but it is surprising that gradualist responses occur almost three times as often (46.5% vs. 16.1%) if a gradualist conclusion is included among a set of predefined answer options. This difference is significant at $p > 0.01$, one-tailed. What explains the difference?

A tentative explanation could be that respondents' willingness to embrace the gradualist hypothesis might be stimulated by a bit of philosophical midwifery. Socrates taught us that philosophers can help people articulate ideas they are unable to express clearly themselves, so this might explain why gradualist answers occur more frequently when gradualist language is included in the vignette. Let us call this the Socratic

https://doi.org/10.1017/9781009336772.003 Published online by Cambridge University Press

midwifery effect. According to this hypothesis, respondents have a tacit but somewhat underdeveloped understanding of the gradualist hypothesis, which is triggered and strengthened by the presence of a clearly stated gradualist answer option, which is consequently selected to a higher extent.

The findings reported in this chapter do not permit us to state any definitive conclusion about the possible significance of a Socratic midwifery effect. About half of the respondents in Study 3 (n = 95) were presented with items C1, C3, and C4. The rest (n = 87) were presented with versions of these items in which the original answer options ("a bit right and a bit wrong," etc.) had been replaced with more complex "midwifery" answer options: "There are reasons for, but also against, doing what [the agent] did. On balance it was neither entirely right nor entirely wrong; it was a bit right but also a bit wrong / it was right / it was wrong / I don't know." For task C1, the percentage of gradualist answers increased from 10.5% to 27.6% (which supports the Socratic midwifery effect), but for C3, the percentage of gradualist answers decreased from 49.5% to 12.6%, and for C4, it decreased from 55.8% to 19.5%. This speaks against the Socratic midwifery effect and there is no obvious explanation of this anomaly. The significance of the Socratic midwifery effect could be a topic for future research.

Objections and Replies

The empirical findings favor the gradualist hypothesis over the binary theory. The abstract, semiabstract, concrete, and comparative measures indicate that if the way in which people use concepts is a reliable guide to meaning, then rightness and wrongness come in degrees. However, no empirical study is immune to criticism.

First, one could worry that other, alternative hypotheses might be equally well supported by data. Consider, for instance, the suggestion that although no act is a bit right *and* a bit wrong, some right acts are right to a greater degree than other right acts, while some wrong acts are more wrong than other wrong acts. According to this alternative hypothesis, RIGHT and WRONG function much like HOT and COLD: some cold objects are colder than other cold objects, but no cold object is a bit cold *and* a bit hot. Data for the comparative tasks do not discriminate between this alternative hypothesis and the gradualist one. All we can conclude from the observation that some items (e.g., items C3–C6) are located "between" an act judged to be entirely wrong (C1)

https://doi.org/10.1017/9781009336772.003 Published online by Cambridge University Press

and others judged to be entirely right (C2 and C7) in a multidimensional scaling is that traditional binary analyses offer a poor fit. The hypothesis that RIGHT and WRONG function like HOT and COLD is, however, compatible with this result.

In response to this, the gradualist could point out that the alternative hypothesis is not compatible with some of the other findings reported here. For instance, the gradualist answer option "right to some degree, but also wrong to some degree" was the most commonly selected response for items C3–C6, that is, the items located "between" acts judged to be entirely right and wrong. This is evidence against the hypothesis that RIGHT and WRONG function like HOT and COLD. If this hypothesis had been true, respondents would not have selected "right to some degree, but also wrong to some degree" to the extent they did.

That said, it is of course possible to formulate other hypotheses that might fit better with data. Suppose, for instance, that respondents believe that acts are complex wholes composed of multivalent parts or aspects. According to this hypothesis, an act could be right with respect to one of its parts or aspects (example: respect for autonomy) but wrong with respect to some other part or aspect (example: fairness), but still be outright right all things considered. If so, respondents who responded "right to some degree, but also wrong to some degree" could mistakenly have selected a gradualist phrase for expressing the view that an act is wrong with respect to some but not all of its parts or aspects.

A drawback of this alternative hypothesis is that it does not square well with some other data points. Consider, for instance, the semiabstract item S1. The vast majority (83.6% in Study 1) reported that lying in a situation in which doing so would bring about the best consequences is "sometimes right to some degree, but also wrong to some degree." If respondents had believed that acts are complex wholes composed of different parts or aspects, they would arguably have selected "I don't know" or "not possible to assess" to a higher extent than they did, as it would have been unclear if the statement referred to the entire act or some of its parts or aspects. However, those answers were selected by no more than 4.4% and 5.8% of respondents. This indicates that the distinction between wholes and parts does not offer a better explanation of data. Another data point that does not fit well with the whole-part hypothesis is the observation that respondents agreed with the abstract statement that "moral rightness comes in degrees" to roughly the same extent that "colors come in degrees." This is also difficult to reconcile with the part-whole hypothesis if we believe that colors come in degrees in an outright sense.

It is worth keeping in mind that *all* theories are underdetermined by data, as emphasized by Quine and others.[10] No matter what evidence we gather for the gradualist hypotheses, it will always be possible to imagine some alternative hypothesis that fits equally well with the experimental findings. We will never be able to prove with certainty that any single hypothesis is the uniquely best one. The modest conclusion of this chapter is, therefore, that the findings reported here offer a better fit with the gradualist hypothesis than any of the alternative hypotheses discussed so far.

This brings us to what is perhaps the most important worry about the meaning-tracks-use argument: Is the fact that ordinary people seem to use RIGHT and WRONG in a sense that permit for degrees a good reason for revising traditional, binary moral theories? If traditional binary moral theories are meant to capture the same concept we use in everyday moral discussions, and meaning tracks use, then the answer is yes. However, a possible response from binary theorists could be that the notions of RIGHT and WRONG described in traditional moral theories are technical concepts, just like many scientific concepts. In light of this, a central task of the next chapter will be to argue that the analogy with technical scientific concepts is problematic. Scientists use technical concepts for a good reason, but there is no analogous good reason to use binary concepts of RIGHT and WRONG in ethics. The technical concept of HEAT in physics enables scientists to express nuanced claims about physical processes that cannot be expressed by the everyday concept. But the binary concepts of RIGHT and WRONG are *less* nuanced than their gradualist counterparts. What would the point be of introducing technical concepts of RIGHT and WRONG that are less sophisticated than our ordinary concepts? By asserting that an act is somewhat right and somewhat wrong we can express information that is lost if we adopt a binary theory that forces us to conclude that every act is either right or wrong simpliciter.

[10] Quine (1975).