

## 551 Appendix

### 552 A CommonsenseQA Error Patterns

553 Throughout our experiments, we came across a variety of interesting failure cases for commonsense  
554 reasoning. Note that all the final answers are correct – however, we take issue with the reasoning  
555 used in order to arrive at those answers.

#### 556 A.1 Question Implies Answer

557 One key failure case was answers in the form of “the answer must be something that is <question  
558 property>. <answer> is <question property>. Therefore, the correct answer is <answer>.” In these  
559 cases, the model fails to explain *why* the answer that it has chosen satisfies the question property.

560 Q: What might someone get from learning about science?

561 Answer Choices:

562 (a) headache

563 (b) see things differently

564 (c) increased knowledge

565 (d) accidents

566 (e) appreciation of nature

567 A: The answer must be something that someone would get from learning about  
568 science. Learning about science would increase knowledge. Therefore, the  
569 answer is increased knowledge (c).

570 These rationales, while perhaps useful to the model, read to us as opaque and unexplanatory.

#### 571 A.2 Begging the Question

572 A related but stronger version of the previous failure case, while less common, is particularly  
573 uninsightful. Sometimes the model will imply the answer that it has chosen in its question.

574 Q: The christian rock band was on a large stage, the singer’s foot got caught  
575 on a wire and he had a small what?

576 Answer Choices:

577 (a) commit sin

578 (b) fail

579 (c) disobey

580 (d) stumble

581 (e) boo boo

582 A: The answer must be something that would cause the singer to stumble. The  
583 singer’s foot got caught on a wire, so he would stumble. Therefore, the  
584 answer is stumble (d).

#### 585 A.3 Exercise to the Reader

586 A rare failure case is when the model finds it unnecessary to justify its answer. For example:

587 Q: Other than school or from text where can one gain knowledge?

588 Answer Choices:

589 (a) book

590 (b) class

591 (c) thinking

592 (d) head

593 (e) meeting

594 A: The answer must be a place where someone can gain knowledge other than  
595 school or from text. The answer is meeting (e).

#### 596 A.4 World State Assertions

597 Sometimes, the model will assume that it knows something about a subject or a person whose name  
598 was used as a variable. This leads to somewhat comical examples of reasoning. Part of the reason  
599 for this is that generally, there is an expectation that good rationales will leverage understanding of  
600 more general classes of objects and appeal to the relationship between those general classes and the  
601 particular instance. For example, the argument that “a person would typically feel exhilaration from  
602 heights” is generally more compelling than the argument that “James would feel exhilaration from  
603 heights.”

604 Q: A person’s hand hurts after hitting the bottle. What might happen to  
605 the bottle?

606 Answer Choices:

- 607 (a) alcoholism
- 608 (b) breaking
- 609 (c) fly
- 610 (d) nausea
- 611 (e) ticket

612 A: The answer must be something that happens to the bottle. The bottle is  
613 broken. Therefore, the answer is breaking (b).

614 Q: James loved climbing. What emotion might he feel from heights?

615 Answer Choices:

- 616 (a) exhaustion
- 617 (b) exhilaration
- 618 (c) falling down
- 619 (d) nervousness
- 620 (e) vertigo

621 A: The answer must be a feeling that would cause someone to climb a mountain.  
622 James would feel exhilaration from heights. Therefore, the answer is  
623 exhilaration (b).

#### 624 A.5 Red Herrings

625 Some errors in reasoning corresponded to the model making a statement which, while technically  
626 true, is not useful in demonstrating the claim.

627 Q: A squirrel ran up to it’s home, where is the small dog likely to be  
628 barking up?

629 Answer Choices:

- 630 (a) manhattan apartment
- 631 (b) wood
- 632 (c) neighbor’s yard
- 633 (d) tree
- 634 (e) basket

635 A: The answer must be a place where a small dog would be likely to be  
636 barking. Squirrels are not small dogs. Therefore, the answer is tree (d).

637 Q: John ran very fast, his feet touched the ground in what way?

638 Answer Choices:

- 639 (a) behind
- 640 (b) loose
- 641 (c) feverishly
- 642 (d) slowly
- 643 (e) lightly

644 A: The answer must be a way of running that is different from the way John  
645 ran. Running is a fast, light-footed activity. Therefore, the answer is  
646 lightly (e).

## 647 **A.6 Hint Short-cutting**

648 In the experiments where the model was shown some examples of “hints” during training, in order to  
649 prevent it from losing the ability to perform rationalization over time, the model appeared to pick up  
650 on the fact that the final answer would always correspond to the hinted answer. This led to answers  
651 such as

652 Q: Billy bought coffee and waited for his wife to arrive from France.  
653 Where might he have been?

654 Answer Choices:

- 655 (a) airport
- 656 (b) grocery store
- 657 (c) internet cafe
- 658 (d) supermarket
- 659 (e) train station

660 A: The answer must be a place where Billy could have been waiting for his  
661 wife to arrive from France. The airport is a place where people can wait  
662 for flights. Therefore, the answer is train station (e).

## 663 B Modified CQA Prompts

664 For reference, we include our modified prompts based closely on those in [6].

665 Q: What do people use to absorb extra ink from a fountain pen?

666 Answer Choices:

- 667 (a) shirt pocket
- 668 (b) calligrapher's hand
- 669 (c) inkwell
- 670 (d) desk drawer
- 671 (e) blotter

672 A: The answer must be used to absorb extra ink. Blotters are designed to absorb  
673 liquids. Therefore, the answer is blotter (e).

674

675 Q: What home entertainment equipment requires cable?

676 Answer Choices:

- 677 (a) radio shack
- 678 (b) substation
- 679 (c) television
- 680 (d) cabinet
- 681 (e) desk

682 A: The answer must require cable. Cable is used to provide satellite channels to  
683 televisions. Therefore, the answer is television (c).

684

685 Q: The fox walked from the city into the forest, what was it looking for?

686 Answer Choices:

- 687 (a) pretty flowers
- 688 (b) hen house
- 689 (c) natural habitat
- 690 (d) storybook
- 691 (e) dense forest

692 A: The answer must be a reason for a fox to go into the forest. The forest is a fox's  
693 natural habitat. Therefore, the answer is natural habitat (c).

694

695 Q: Sammy wanted to go to where the people were. Where might he go?

696 Answer Choices:

- 697 (a) populated areas
- 698 (b) race track
- 699 (c) desert
- 700 (d) apartment
- 701 (e) roadblock

702 A: The answer must be a place with many people. Populated areas, by definition, have  
703 a lot of people. Therefore, the answer is populated areas (a).

704

705 Q: Where do you put your grapes just before checking out?

706 Answer Choices:

- 707 (a) mouth
- 708 (b) grocery cart
- 709 (c) super market
- 710 (d) fruit basket
- 711 (e) fruit market

712 A: The answer should be the place where grocery items are placed before checking out  
713 . Of the above choices, grocery cart makes the most sense for holding grocery items.  
714 Therefore, the answer is grocery cart (b).

715

716 Q: Google Maps and other highway and street GPS services have replaced what?

717 Answer Choices:

- 718 (a) united states
- 719 (b) mexico
- 720 (c) countryside
- 721 (d) atlas
- 722 (e) oceans

723 A: The answer must be something that used to do what Google Maps and GPS services do  
724 , which is give directions. Atlases were also used to give directions. Therefore,  
725 the answer is atlas (d).

726  
727 Q: Before getting a divorce, what did the wife feel who was doing all the work?  
728 Answer Choices:  
729 (a) harder  
730 (b) anguish  
731 (c) bitterness  
732 (d) tears  
733 (e) sadness  
734 A: The answer should be a feeling which would cause someone who was doing all the  
735 work to get divorced. If someone feels bitter towards their spouse, they are likely  
736 to want a divorce. Therefore, the answer is bitterness (c).

## 737 C Human-evaluated Test Prompts

738 We also selected a random sampling of 50 questions which were correctly answered both few-shot  
 739 and by a STaR-trained model (without rationalization), as discussed in 4.4. Presented in a random  
 740 order, twenty crowdworkers preferred the STaR-generated answers. We reproduce the examples here  
 741 with the few-shot rationale first, the STaR-trained rationale second, and the human rationale third,  
 742 though these were shuffled when presented to participants. We selected human answers from [3]’s  
 743 original split rationales where possible, finding that duplicate rationales were much more common in  
 744 the new split rationales. For example, the explanation “Rivers flow trough valleys,” appeared over  
 745 400 times verbatim in the new split dataset, and “This word was most relevant” appeared over 150  
 746 times. [3]’s dataset also includes explanations like “The only answer that makes sense” or “BOB  
 747 WILL NOT POKEMON CARDS WERE COMMON AND WORTHLESS BUT WRONG ABOUT  
 748 THEM SO FEEL REALLY RARE TO DELAY” or restatements of the answer. We append the phrase  
 749 “Therefore, the answer is ANSWERTEXT (ANSWERLETTER)” with ANSWERTEXT replaced  
 750 by the correct answer’s text and ANSWERLETTER replaced by the correct answer letter. This is  
 751 done 1) to make it less obvious that one of the answers is generated by a different source and 2) to  
 752 prioritize differences in rationales, not the answer format.

753 Before the questions and after the consent form, we presented the following directions:

754  
 755 Each question will have three explanations for one answer, which is a pre-determined correct ("best") answer.  
 756  
 757 Please rank the explanations for each question from best to worst (with 1 as the best and 3 as the worst) based on how well  
 758 they justify the correct answer.  
 759  
 760 Note there are some typos and grammatical errors in the examples, in the questions, answers, and the explanation. Please  
 761 treat the answer letter (a / b / c / d / e) at the end of the explanations as the correct answer letter, even if another  
 762 answer is valid (and perhaps more reasonable), and even if it does not match with the answer text preceding it.

763 The examples were subsampled and presented to the crowdworkers:

764 Q: Who is likely to have a caring heart?	811 A: Having fun is enjoyable. Therefore, the answer is
765 Answer Choices:	812 enjoyable (e).
766 (a) person	813
767 (b) species	814
768 (c) artichoke	815
769 (d) all mammals	816 Q: What happens when I make a pig of myself eating breakfast
770 (e) live animal	817 ?
771 A: The answer must be a person. All mammals have a caring	818 Answer Choices:
772 heart. Therefore, the answer is person (a).	819 (a) full stomach
773 A: The answer must be someone who is likely to have a	820 (b) gain energy
774 caring heart. Humans are the most likely to have a caring	821 (c) dieting
775 heart. Therefore, the answer is person (a).	822 (d) have energy
776 A: A person is the only human being capable of having a	823 (e) feeling satisfied
777 caring heart. Therefore, the answer is person (a).	824
778	825 A: The answer must be a feeling which would cause someone
779 Q: If a person wanted to become a good writer what should	826 to feel full after eating breakfast. The answer is full
780 they do in college?	827 stomach (a).
781 Answer Choices:	828 A: The answer must be a feeling which would cause someone
782 (a) word sentence	829 to make a pig of themselves eating breakfast. If someone
783 (b) own animal	830 feels full, they are likely to stop eating. Therefore, the
784 (c) read the newspaper	831 answer is full stomach (a).
785 (d) catch cold	832 A: A colloquial phrase for over indulging in food.
786 (e) study literature	833 Therefore, the answer is full stomach (a).
787 A: The answer must be something that would help a person	834
788 become a good writer. The answer is study literature (e).	835 Q: In order to see a story on the big screen what must you
789 A: The answer must be something that would help a person	836 do?
790 become a good writer. Literature is a form of study.	837 Answer Choices:
791 Therefore, the answer is study literature (e).	838 (a) go to movies
792 A: studying literature will give one skills of writing.	839 (b) visualize
793 Therefore, the answer is study literature (e).	840 (c) reading
794	841 (d) open book
795 Q: The human always would have fun making up questions for	842 (e) sketching a picture
796 the A.I. overlords, he found the task quite what?	843
797 Answer Choices:	844 A: The answer must be something that you do to see a story
798 (a) do enjoy	845 on the big screen. Movies are a way to see a story on the
799 (b) eat cake	846 big screen. Therefore, the answer is movies (a).
800 (c) enjoy living	847 A: The answer must be something that is required to see a
801 (d) get laid	848 story on the big screen. Movies are a form of storytelling.
802 (e) enjoyable	849 Therefore, the answer is go to movies (a).
803	850 A: The big screen is a colloquial term meaning a movie
804 A: The answer must be something that the human would enjoy	851 theater. Therefore, the answer is go to movies (a).
805 doing. The human enjoys making up questions for the A.I.	852
806 overlords. Therefore, the answer is enjoyable (e).	853 Q: What is the opposite of remembering?
807 A: The answer must be something that the human found	854 Answer Choices:
808 enjoyable. The human found the task of making up questions	855 (a) answering question
809 for the A.I. overlords enjoyable. Therefore, the answer is	856 (b) stopping to think
810 enjoyable (e).	857 (c) knowing
	858 (d) closing eyes

859 (e) forgetting 950  
 860 A: The answer must be something that is the opposite of 951  
 861 remembering. The answer is forgetting (e). 952  
 862 A: The answer must be a feeling which would cause someone 953  
 863 to not remember something. If someone forgets something, 954  
 864 they are likely to not remember it. Therefore, the answer 955  
 865 is forgetting (e). 956  
 866 A: Health complications. Therefore, the answer is 957  
 867 forgetting (e). 958  
 868 959  
 869 Q: If you aren't experiencing curiosity and want to ensure 960  
 870 viewing enjoyment, what may you watch? 961  
 871 Answer Choices: 962  
 872 (a) examine thing 963  
 873 (b) stare into a mirror 964  
 874 (c) go somewhere 965  
 875 (d) hear news 966  
 876 (e) see favorite show 967  
 877 A: The answer must be something that will make you want to 968  
 878 watch. If you are not experiencing curiosity, you may want 969  
 879 to watch something that will make you want to watch. 970  
 880 Therefore, the answer is watch (e). 971  
 881 A: The answer must be something that will not cause you to 972  
 882 want to watch. If you are not experiencing curiosity, you 973  
 883 are likely to want to watch something that is entertaining. 974  
 884 Therefore, the answer is favorite show (e). 975  
 885 A: see favorite show can you do in front of a tv if you 976  
 886 have curiosity about what is being displayed on the tv. 977  
 887 Therefore, the answer is favorite show (e). 978  
 888 979  
 889 Q: The carotid is a main artery susceptible to damage where 980  
 890 Answer Choices: 981  
 891 (a) neck 982  
 892 (b) face 983  
 893 (c) busy city 984  
 894 (d) torso 985  
 895 (e) human body 986  
 896 A: The answer must be a place where the carotid artery is 987  
 897 susceptible to damage. The carotid artery is located in the 988  
 898 neck. Therefore, the answer is neck (a). 989  
 899 A: The answer must be a place where the carotid is 990  
 900 susceptible to damage. The carotid is susceptible to damage 991  
 901 in the neck. Therefore, the answer is neck (a). 992  
 902 A: The carotid artery is located in the neck. Therefore, 993  
 903 the answer is neck (a). 994  
 904 995  
 905 Q: The king needed to feel safe, where did he go? 996  
 906 Answer Choices: 997  
 907 (a) castle 998  
 908 (b) throne room 999  
 909 (c) deck of cards 1000  
 910 (d) fort 1001  
 911 (e) court 1002  
 912 A: The answer must be a place where the king feels safe. 1003  
 913 The king's castle is a place where he feels safe. Therefore, 1004  
 914 the answer is castle (a). 1005  
 915 A: The answer must be a place where the king can feel safe. 1006  
 916 Castles are places where kings feel safe. Therefore, the 1007  
 917 answer is castle (a). 1008  
 918 A: In general, King live in castle. Therefore, the answer 1009  
 919 is castle (a). 1010  
 920 1011  
 921 Q: What is the result of instituting civil action? 1012  
 922 Answer Choices: 1013  
 923 (a) bad feelings 1014  
 924 (b) confusion 1015  
 925 (c) going to court 1016  
 926 (d) hurt feelings 1017  
 927 (e) sandwiches 1018  
 928 A: The answer must be a result of instituting civil action. 1019  
 929 The result of instituting civil action is a lawsuit. 1020  
 930 Therefore, the answer is going to court (c). 1021  
 931 A: The answer must be a result of instituting civil action. 1022  
 932 If someone goes to court, they are likely to feel hurt 1023  
 933 feelings. Therefore, the answer is going to court (c). 1024  
 934 A: Anger and Trial will not institute a court case. 1025  
 935 Therefore, the answer is going to court (c). 1026  
 1027  
 936 Q: The person wanted to fix his car, what did he order? 1028  
 937 Answer Choices: 1029  
 938 (a) selling things 1030  
 939 (b) gas 1031  
 940 (c) feel resentful 1032  
 941 (d) machine part 1033  
 942 (e) cross street 1034  
 943 A: The answer must be something that would help the person 1035  
 944 fix his car. The person would not want to sell things, 1036  
 945 because he would not want to give up his car. Therefore, 1037  
 946 the answer is machine part (d). 1038  
 947 A: The answer must be something that would be used to fix 1039  
 948 car. Machine parts are used to fix cars. Therefore, the 1040  
 949 answer is machine part (d). 1041

A: A machine part is the only thing listed that would fix the car. Therefore, the answer is machine part (d).  
 Q: He always traveled up the mountain when winter arrived, he had a passion and desire to do what?  
 Answer Choices:  
 (a) play volleyball  
 (b) play hockey  
 (c) ski  
 (d) knit  
 (e) skate  
 A: The answer must be something that the person does when winter arrives. The answer is skiing (c).  
 A: The answer must be something that the person did when winter arrived. The answer is ski (c).  
 A: Skiing is the only option that is done on a mountain. Therefore, the answer is ski (c).  
 Q: Bob thought that his Pokemon cards were common and worthless, but he was wrong about them. They were really what?  
 Answer Choices:  
 (a) personal  
 (b) special  
 (c) shiny  
 (d) rare  
 (e) irregular  
 A: The answer must be something that is not common and worthless. Pokemon cards are not common and worthless. Therefore, the answer is rare (d).  
 A: The answer must be something that are rare. Pokemon cards are rare. Therefore, the answer is rare (d).  
 A: BOB WILL NOT POKEMON CARDS WERE COMMON AND WORTHLESS BUT WRONG ABOUT THEM SO FEEL REALLY RARE TO DELAY. Therefore, the answer is rare (d).  
 Q: If you're lighting a candle to see what are you in?  
 Answer Choices:  
 (a) dimly lit room  
 (b) synagogue  
 (c) sunlight  
 (d) birthday cake  
 (e) bedroom  
 A: The answer must be a place where you can see what you are in. The answer is a room with dim lighting. Therefore, the answer is dimly lit room (a).  
 A: The answer must be a place where you can see a candle. Candles are used to see things in the dark. Therefore, the answer is dimly lit room (a).  
 A: Lighting or illumination is the deliberate use of light to achieve a practical or aesthetic effect. Lighting includes the use of both artificial light sources like lamps ... Therefore, the answer is dimly lit room (a).  
 Q: The person was thirsty, what should he have done?  
 Answer Choices:  
 (a) eat  
 (b) sunny days  
 (c) compliments  
 (d) drink water  
 (e) all kinds of things  
 A: The answer must be something that would make the person feel better. If the person is thirsty, he should drink water. Therefore, the answer is drink water (d).  
 A: The answer must be something that would make someone feel better if they were thirsty. If someone is thirsty, they are likely to want to drink water. Therefore, the answer is drink water (d).  
 A: The person was thirsty, so drink water. Therefore, the answer is drink water (d).  
 Q: Where would you find many products for sale?  
 Answer Choices:  
 (a) market  
 (b) cost one dollar  
 (c) shelf  
 (d) salon  
 (e) demonstration  
 A: The answer must be a place where many products are sold. A market is a place where many products are sold. Therefore, the answer is market (a).  
 A: The answer must be a place with a lot of products for sale. Markets are the most common place to buy products. Therefore, the answer is market (a).  
 A: The only place products are for sale are in markets and stores. Therefore, the answer is market (a).  
 Q: If man and mammoth wanted to meet, they would probably have to do it during when?  
 Answer Choices:  
 (a) a baby shower  
 (b) ancient history

1042 (c) storybook 1133  
 1043 (d) stone age 1134  
 1044 (e) forest 1135  
 1045 A: The answer must be a time when man and mammoth would 1136  
 1046 meet. Stone age is the time when man and mammoth lived. 1137  
 1047 Therefore, the answer is stone age (d). 1138  
 1048 A: The answer must be a time when man and mammoth would 1139  
 1049 able to meet. The stone age was a time when man and mammoth 1140  
 1050 would be able to meet. Therefore, the answer is stone age 1141  
 1051 (d). 1142  
 1052 A: During the stone age is the only time mammoths lived. 1143  
 1053 Therefore, the answer is stone age (d). 1144  
 1054 1145  
 1055 Q: They prided themselves on being a wealth of knowledge, 1146  
 1056 and that's why many chose to attend their what? 1147  
 1057 Answer Choices: 1148  
 1058 (a) book store 1149  
 1059 (b) university 1150  
 1060 (c) meeting 1151  
 1061 (d) class 1152  
 1062 (e) encyclopedia 1153  
 1063 A: The answer must be a place where people go to learn 1154  
 1064 about things. The answer is university (b). 1155  
 1065 A: The answer must be a place where people go to learn 1156  
 1066 about things. Universities are places where people go to 1157  
 1067 learn about things. Therefore, the answer is university 1158  
 1068 1159  
 1069 A: because in the university you get more knowledge. 1160  
 1070 Therefore, the answer is university (b). 1161  
 1071 1162  
 1072 Q: what happens when visiting museum that have narrow hall 1163  
 1073 Answer Choices: 1164  
 1074 (a) sore feet 1165  
 1075 (b) amazement 1166  
 1076 (c) retrospection 1167  
 1077 (d) boredom 1168  
 1078 (e) crowding 1169  
 1079 A: The answer must be something that happens when visiting 1170  
 1080 a museum that has narrow halls. Crowding is a common 1171  
 1081 problem in museums. Therefore, the answer is crowding (e) 1172  
 1082 A: The answer must be something that would cause someone 1173  
 1083 have sore feet if they are walking in narrow halls. If 1174  
 1084 someone feels cramped, they are likely to want to go 1175  
 1085 somewhere else. Therefore, the answer is crowding (e). 1176  
 1086 A: museum is one of the crowdest place. Therefore, the 1177  
 1087 answer is crowding (e). 1178  
 1088 1179  
 1089 Q: On their hike they brought a filtering straw, they were 1180  
 1090 worried about germs in the what? 1181  
 1091 Answer Choices: 1182  
 1092 (a) make sick 1183  
 1093 (b) doctor 1184  
 1094 (c) water 1185  
 1095 (d) stream 1186  
 1096 (e) mouth 1187  
 1097 A: The answer must be something that can filter out germs 1188  
 1098 Filtering straws are used to filter out germs. Therefore 1189  
 1099 the answer is filtering straw (c). 1190  
 1100 A: The answer must be something that would cause someone 1191  
 1101 bring a filtering straw on a hike. Filtering straws are 1192  
 1102 used to filter water. Therefore, the answer is water (c) 1193  
 1103 A: straw it something used to drink water. Therefore, the 1194  
 1104 answer is water (c). 1195  
 1105 1196  
 1106 Q: If I am not good at playing poker, I will begin what 1197  
 1107 play? 1198  
 1108 Answer Choices: 1199  
 1109 (a) discord 1200  
 1110 (b) social event 1201  
 1111 (c) losing money 1202  
 1112 (d) competition 1203  
 1113 (e) heightened awareness 1204  
 1114 A: The answer must be something that will happen if you 1205  
 1115 not good at playing poker. If you are not good at playing 1206  
 1116 poker, you will begin to lose money. Therefore, the answer 1207  
 1117 is losing money (c). 1208  
 1118 A: The answer must be something that will help you if you 1209  
 1119 are not good at playing poker. If you are losing money, 1210  
 1120 are likely to want to stop playing. Therefore, the answer 1211  
 1121 is losing money (c). 1212  
 1122 A: Poker is a game in which people lose their money. 1213  
 1123 Therefore, the answer is losing money (c). 1214  
 1124 1215  
 1125 Q: People are meeting people, what is the likely physical 1216  
 1126 gesture? 1217  
 1127 Answer Choices: 1218  
 1128 (a) introducing yourself 1219  
 1129 (b) smiling 1220  
 1130 (c) shake hands 1221  
 1131 (d) share a hug 1222  
 1132 (e) talking 1223  
 A: The answer must be a physical gesture that is used to  
 greet someone. The answer is handshake (c).  
 A: The answer must be a gesture used to meet people. If  
 someone wants to meet people, they are likely to want to  
 shake hands. Therefore, the answer is shaking hands (c).  
 A: A handshake is a globally widespread. Therefore, the  
 answer is shaking hands (c).  
 Q: What do dogs get that make them itch?  
 Answer Choices:  
 (a) bones  
 (b) fleas  
 (c) bad breath  
 (d) wagging tails  
 (e) four legs  
 A: The answer must be something that makes dogs itch. Dogs  
 get fleas, which make them itch. Therefore, the answer is  
 fleas (b).  
 A: The answer must be something that dogs get that make  
 them itch. Fleas are a common cause of dog itch. Therefore,  
 the answer is fleas (b).  
 A: fleas makes you itch. Therefore, the answer is fleas (b).  
 Q: What is ignorance often equated with?  
 Answer Choices:  
 (a) bliss  
 (b) hostility  
 (c) peace  
 (d) accidents  
 (e) find truth  
 A: The answer must be something that is often equated with  
 ignorance. Bliss is often equated with ignorance. Therefore,  
 the answer is bliss (a).  
 A: The answer must be a feeling which would cause someone  
 to be ignorant. If someone feels bliss, they are likely to  
 be ignorant. Therefore, the answer is bliss (a).  
 A: bliss comes from ignorance. Therefore, the answer is  
 bliss (a).  
 Q: The two teams were competing all the time, what likely  
 developed between them?  
 Answer Choices:  
 (a) accomplishment  
 (b) enemies  
 (c) rivalry  
 (d) friendship  
 (e) bad blood  
 A: The answer must be something that developed between the  
 two teams. The teams were competing all the time. Therefore,  
 the answer is rivalry (c).  
 A: The answer must be a feeling which would cause two teams  
 to compete all the time. Rivalry is a common feeling which  
 would cause two teams to compete all the time. Therefore,  
 the answer is rivalry (c).  
 A: likely developed between them of rivalry. Therefore, the  
 answer is rivalry (c).  
 Q: James doesn't want a relationship. He thinks that loving  
 another will only create what?  
 Answer Choices:  
 (a) satisfaction  
 (b) jealousy  
 (c) happiness  
 (d) wonderful  
 (e) quiet  
 A: The answer must be something that would make James feel  
 bad about not having a relationship. If someone feels  
 jealous, they are likely to want a relationship. Therefore,  
 the answer is jealousy (b).  
 A: The answer must be a feeling which would cause someone  
 who doesn't want a relationship to not want one. If someone  
 feels jealous of another, they are likely to not want a  
 relationship. Therefore, the answer is jealousy (b).  
 A: Jealousy is a negative emotion. Therefore, the answer is  
 jealousy (b).  
 Q: The had a reputation for celebrating in a debaucherous  
 manner, what often happened as a result?  
 Answer Choices:  
 (a) drunkenness  
 (b) enjoyment  
 (c) have fun  
 (d) hang over  
 (e) fighting  
 A: The answer must be a result of a celebration. The answer  
 is drunkenness (a).  
 A: The answer must be something that would cause someone to  
 have a reputation for celebrating in a debaucherous manner.  
 Alcohol is a depressant. Therefore, the answer is  
 drunkenness (a).

1224 A: it is normal not a good idea to get drunk. Therefore, the answer is drunkenness (a).  
1225  
1226  
1227 Q: The paper towels are usually in plain sight in the kitchen, where did the dad go next to find them?  
1228 Answer Choices:  
1229 (a) pantry  
1230 (b) cabinet  
1231 (c) grocery store  
1232 (d) waste bin  
1233 (e) market  
1234 A: The answer must be a place where the dad would go to find the paper towels. The pantry is a place where food is stored. Therefore, the answer is pantry (a).  
1235 A: The answer must be a place where paper towels are stored. Pantries are the most likely place to store paper towels. Therefore, the answer is pantry (a).  
1236 A: The pantry is in the kitchen and stores things. Therefore, the answer is pantry (a).  
1237  
1238 Q: My house is very dry, with little water in the air. What tool should I use to fix this?  
1239 Answer Choices:  
1240 (a) humidifier  
1241 (b) surface of earth  
1242 (c) skating rink  
1243 (d) ice cube  
1244 (e) burn it to the ground  
1245 A: The answer must be a tool that will make the air in your house moist. A humidifier is a tool that will make the air in your house moist. Therefore, the answer is humidifier (a).  
1246 A: The answer must be something that would make the air in my house more humid. A humidifier is used to make the air in a room more humid. Therefore, the answer is humidifier (a).  
1247 A: This word was most relevant. Therefore, the answer is humidifier (a).  
1248  
1249 Q: A motorcycle is quicker because it isn't as what as a car?  
1250 Answer Choices:  
1251 (a) heavy  
1252 (b) shiny  
1253 (c) red  
1254 (d) obesity  
1255 (e) expensive  
1256 A: The answer must be something that is not as heavy as a car. A motorcycle is lighter than a car. Therefore, the answer is not heavy (a).  
1257 A: The answer must be something that is not as quick as a motorcycle. Motorcycles are not as heavy as cars. Therefore, the answer is heavy (a).  
1258 A: Because it is much smaller. Therefore, the answer is heavy (a).  
1259  
1260 Q: The baseball team was playing ball. It was time for the pitcher to act. How did the pitcher get the ball to the batter?  
1261 Answer Choices:  
1262 (a) lose  
1263 (b) injury  
1264 (c) competition  
1265 (d) throwing  
1266 (e) rolled  
1267 A: The answer must be a way to get the ball to the batter. The pitcher is the one who throws the ball. Therefore, the answer is throwing (d).  
1268 A: The answer must be a way for the pitcher to get the ball to the batter. The pitcher is the person who is responsible for getting the ball to the batter. Therefore, the answer is throwing (d).  
1269 A: The pitcher threw the ball to the batter so that he could hit it. Therefore, the answer is throwing (d).  
1270  
1271 Q: In what sort of container should you keep your ficus?  
1272 Answer Choices:  
1273 (a) refrigerator  
1274 (b) screened porch  
1275 (c) clay pot  
1276 (d) shady places  
1277 (e) arboretum  
1278 A: The answer must be a container that keeps the ficus in a shady place. Clay pots are used to keep plants in shady places. Therefore, the answer is clay pot (c).  
1279 A: The answer must be a container in which ficus can be kept. Ficus are tropical plants. Therefore, the answer is clay pot (c).  
1280 A: Clay pot is the only type of container listed. Therefore, the answer is clay pot (c).  
1281  
1282 Q: James was cooking s stew in his apartment kitchen. Where might he look for a potato.  
1283 Answer Choices:  
1284 (a) restaurants  
1285 (b) casserole  
1286 (c) root cellar  
1287 (d) pockets  
1288 (e) pantry  
1289 A: The answer must be a place where potatoes are stored. The answer is pantry (e).  
1290 A: The answer must be a place where potatoes are stored. Pantries are the most likely place to store potatoes. Therefore, the answer is pantry (e).  
1291 A: apartments do not often have root cellars, and are not restaurants. a pantry is closest to James inside his apartment. Therefore, the answer is pantry (e).  
1292  
1293 Q: If ships are in a storm and the sun comes out, what has happened?  
1294 Answer Choices:  
1295 (a) sinks  
1296 (b) cross river  
1297 (c) collide  
1298 (d) bottom out  
1299 (e) carry people  
1300 A: The answer must be a thing that happens when ships are in a storm and the sun comes out. Ships are in a storm when they are in the middle of the ocean. Therefore, the answer is bottom out (d).  
1301 A: The answer must be something that happens when the sun comes out. The sun is a source of energy. Therefore, the answer is bottom out (d).  
1302 A: the sun comes out so its not a storm any more. Therefore, the answer is bottom out (d).  
1303  
1304 Q: Where do all marmot live before they are captured?  
1305 Answer Choices:  
1306 (a) switzerland  
1307 (b) spain  
1308 (c) vancouver island  
1309 (d) wilderness  
1310 (e) australia  
1311 A: The answer must be a place where marmots live. Marmots live in the wilderness. Therefore, the answer is wilderness (d).  
1312 A: The answer must be a place where marmots are found. The wilderness is the most likely place to find marmots. Therefore, the answer is wilderness (d).  
1313 A: Marmots don't live in Australia or Switzerland. Therefore, the answer is wilderness (d).  
1314  
1315 Q: Where does one wait on a bench for something on tracks?  
1316 Answer Choices:  
1317 (a) train station  
1318 (b) state park  
1319 (c) crosswalk  
1320 (d) bus stop  
1321 (e) bus depot  
1322 A: The answer must be a place where one waits for a train, bus, or other vehicle. The answer is train station (a).  
1323 A: The answer must be a place where one can wait for a train. Train stations are places where trains stop. Therefore, the answer is train station (a).  
1324 A: Only trains are on tracks. Therefore, the answer is train station (a).  
1325  
1326 Q: What can planning vacation lead to between a couple when they disagree?  
1327 Answer Choices:  
1328 (a) relaxation  
1329 (b) enjoying  
1330 (c) arguments  
1331 (d) going abroad  
1332 (e) spending money  
1333 A: The answer must be something that can happen when a couple disagrees. Arguments are a common result of disagreements. Therefore, the answer is arguments (c).  
1334 A: The answer must be a result of planning vacation. Arguments are a common result of planning vacation. Therefore, the answer is arguments (c).  
1335 A: Arguments are the direct result of disagreements. Therefore, the answer is arguments (c).  
1336  
1337 Q: A sloth climbs slow, but it can eventually reach all the way up where?  
1338 Answer Choices:  
1339 (a) forest canopy  
1340 (b) universe  
1341 (c) tropical rainforest  
1342 (d) large trees  
1343 (e) nature preserve

1406 A: The answer must be a place where a sloth can climb. Therefore, the answer is forest canopy (a).  
1407  
1408  
1409 A: The answer must be a place where a sloth can climb. Therefore, the answer is forest canopy (a).  
1410  
1411  
1412 A: The sloth is a mammal that lives in the trees of South America. They have a slow and peculiar way of moving about. In fact, they are the slowest land mammals and travel sluggishly on the ground (0.025 m/s-0.050 m/s). Therefore, the answer is canopy of a forest (a).  
1413  
1414  
1415  
1416  
1417  
1418 Q: Where in America is known for theaters?  
1419 Answer Choices:  
1420 (a) new york city  
1421 (b) shopping mall  
1422 (c) downtown area  
1423 (d) skating rink  
1424 (e) military base  
1425  
1426 A: The answer must be a place where people go to see movies. Theaters are places where people go to see movies. Therefore, the answer is theater (a).  
1427  
1428  
1429 A: The answer must be a place with a lot of theaters. New York City is known for its many theaters. Therefore, the answer is New York City (a).  
1430  
1431  
1432 A: New York City is the main region in US known for theaters. Therefore, the answer is New York City (a).  
1433  
1434  
1435 Q: What does hearing someone play violin beautifully make you feel?  
1436 Answer Choices:  
1437 (a) buy earplugs  
1438 (b) inspiring  
1439 (c) guitar  
1440 (d) make music  
1441 (e) like music  
1442  
1443 A: The answer must be something that makes you feel inspired. Hearing someone play violin beautifully makes me feel inspired. Therefore, the answer is inspiring (b).  
1444  
1445  
1446 A: The answer must be a feeling which would cause someone to want to hear someone play violin beautifully. Music is the form of art. Therefore, the answer is inspiring (b).  
1447  
1448  
1449 A: Hearing someone play a violin beautifully makes me inspired. Therefore, the answer is inspiring (b).  
1450  
1451  
1452 Q: What emotion does committing murder induce?  
1453 Answer Choices:  
1454 (a) fear  
1455 (b) go to jail  
1456 (c) problems  
1457 (d) cruel  
1458 (e) dead  
1459  
1460 A: The answer must be a feeling that would cause someone to commit murder. Murder is a crime. Therefore, the answer is fear (a).  
1461  
1462  
1463 A: The answer must be a feeling which would cause someone to commit murder. If someone feels fear, they are likely to want to avoid committing murder. Therefore, the answer is fear (a).  
1464  
1465  
1466 A: The only feeling. Therefore, the answer is fear (a).  
1467  
1468 Q: Sitting too close while watching TV can cause what sort of pain?  
1469 Answer Choices:  
1470 (a) brain problems  
1471 (b) laziness  
1472 (c) get fat  
1473 (d) headache  
1474 (e) laughter  
1475  
1476 A: The answer must be a pain that comes from sitting too close to the TV. Sitting too close to the TV can cause headaches. Therefore, the answer is headache (d).  
1477  
1478  
1479 A: The answer must be a type of pain which can occur if you are sitting too close to the TV. Headaches are a common side effect of sitting too close to the TV. Therefore, the answer is headache (d).  
1480  
1481  
1482  
1483 A: the process of aging in humans. Therefore, the answer is headache (d).  
1484  
1485  
1486 Q: What might people do to protect their legs from getting dirty on the farm?  
1487 Answer Choices:  
1488 (a) wear jeans  
1489 (b) milk cow  
1490 (c) believe in god  
1491 (d) avoid mud  
1492 (e) plant flowers  
1493  
1494 A: The answer must be something that people do to protect their legs from getting dirty. Of the above choices, wearing jeans is the most logical. Therefore, the answer is jeans (a).  
1495  
1496  
1497  
1498  
1499  
1500  
1501  
1502  
1503  
1504  
1505  
1506  
1507  
1508  
1509  
1510  
1511  
1512  
1513  
1514  
1515  
1516  
1517  
1518  
1519  
1520  
1521  
1522  
1523  
1524  
1525  
1526  
1527  
1528  
1529  
1530  
1531  
1532  
1533  
1534  
1535  
1536  
1537  
1538  
1539  
1540  
1541  
1542  
1543  
1544  
1545  
1546  
1547  
1548  
1549  
1550  
1551  
1552  
1553  
1554  
1555  
1556  
1557  
1558  
1559  
1560  
1561  
1562  
1563  
1564  
1565  
1566  
1567  
1568  
1569  
1570  
1571  
1572  
1573  
1574  
1575  
1576  
1577  
1578  
1579  
1580  
1581  
1582  
1583  
1584  
1585  
1586  
1587  
1588  
1589  
1590  
1591  
1592  
1593  
1594  
1595  
1596  
1597  
1598  
1599  
1600  
1601  
1602  
1603  
1604  
1605  
1606  
1607  
1608  
1609  
1610  
1611  
1612  
1613  
1614  
1615  
1616  
1617  
1618  
1619  
1620  
1621  
1622  
1623  
1624  
1625  
1626  
1627  
1628  
1629  
1630  
1631  
1632  
1633  
1634  
1635  
1636  
1637  
1638  
1639  
1640  
1641  
1642  
1643  
1644  
1645  
1646  
1647  
1648  
1649  
1650  
1651  
1652  
1653  
1654  
1655  
1656  
1657  
1658  
1659  
1660  
1661  
1662  
1663  
1664  
1665  
1666  
1667  
1668  
1669  
1670  
1671  
1672  
1673  
1674  
1675  
1676  
1677  
1678  
1679  
1680  
1681  
1682  
1683  
1684  
1685  
1686  
1687  
1688  
1689  
1690  
1691  
1692  
1693  
1694  
1695  
1696  
1697  
1698  
1699  
1700  
1701  
1702  
1703  
1704  
1705  
1706  
1707  
1708  
1709  
1710  
1711  
1712  
1713  
1714  
1715  
1716  
1717  
1718  
1719  
1720  
1721  
1722  
1723  
1724  
1725  
1726  
1727  
1728  
1729  
1730  
1731  
1732  
1733  
1734  
1735  
1736  
1737  
1738  
1739  
1740  
1741  
1742  
1743  
1744  
1745  
1746  
1747  
1748  
1749  
1750  
1751  
1752  
1753  
1754  
1755  
1756  
1757  
1758  
1759  
1760  
1761  
1762  
1763  
1764  
1765  
1766  
1767  
1768  
1769  
1770  
1771  
1772  
1773  
1774  
1775  
1776  
1777  
1778  
1779  
1780  
1781  
1782  
1783  
1784  
1785  
1786  
1787  
1788  
1789  
1790  
1791  
1792  
1793  
1794  
1795  
1796  
1797  
1798  
1799  
1800  
1801  
1802  
1803  
1804  
1805  
1806  
1807  
1808  
1809  
1810  
1811  
1812  
1813  
1814  
1815  
1816  
1817  
1818  
1819  
1820  
1821  
1822  
1823  
1824  
1825  
1826  
1827  
1828  
1829  
1830  
1831  
1832  
1833  
1834  
1835  
1836  
1837  
1838  
1839  
1840  
1841  
1842  
1843  
1844  
1845  
1846  
1847  
1848  
1849  
1850  
1851  
1852  
1853  
1854  
1855  
1856  
1857  
1858  
1859  
1860  
1861  
1862  
1863  
1864  
1865  
1866  
1867  
1868  
1869  
1870  
1871  
1872  
1873  
1874  
1875  
1876  
1877  
1878  
1879  
1880  
1881  
1882  
1883  
1884  
1885  
1886  
1887  
1888  
1889  
1890  
1891  
1892  
1893  
1894  
1895  
1896  
1897  
1898  
1899  
1900  
1901  
1902  
1903  
1904  
1905  
1906  
1907  
1908  
1909  
1910  
1911  
1912  
1913  
1914  
1915  
1916  
1917  
1918  
1919  
1920  
1921  
1922  
1923  
1924  
1925  
1926  
1927  
1928  
1929  
1930  
1931  
1932  
1933  
1934  
1935  
1936  
1937  
1938  
1939  
1940  
1941  
1942  
1943  
1944  
1945  
1946  
1947  
1948  
1949  
1950  
1951  
1952  
1953  
1954  
1955  
1956  
1957  
1958  
1959  
1960  
1961  
1962  
1963  
1964  
1965  
1966  
1967  
1968  
1969  
1970  
1971  
1972  
1973  
1974  
1975  
1976  
1977  
1978  
1979  
1980  
1981  
1982  
1983  
1984  
1985  
1986  
1987  
1988  
1989  
1990  
1991  
1992  
1993  
1994  
1995  
1996  
1997  
1998  
1999  
2000  
2001  
2002  
2003  
2004  
2005  
2006  
2007  
2008  
2009  
2010  
2011  
2012  
2013  
2014  
2015  
2016  
2017  
2018  
2019  
2020  
2021  
2022  
2023  
2024  
2025  
2026  
2027  
2028  
2029  
2030  
2031  
2032  
2033  
2034  
2035  
2036  
2037  
2038  
2039  
2040  
2041  
2042  
2043  
2044  
2045  
2046  
2047  
2048  
2049  
2050  
2051  
2052  
2053  
2054  
2055  
2056  
2057  
2058  
2059  
2060  
2061  
2062  
2063  
2064  
2065  
2066  
2067  
2068  
2069  
2070  
2071  
2072  
2073  
2074  
2075  
2076  
2077  
2078  
2079  
2080  
2081  
2082  
2083  
2084  
2085  
2086  
2087  
2088  
2089  
2090  
2091  
2092  
2093  
2094  
2095  
2096  
2097  
2098  
2099  
2100  
2101  
2102  
2103  
2104  
2105  
2106  
2107  
2108  
2109  
2110  
2111  
2112  
2113  
2114  
2115  
2116  
2117  
2118  
2119  
2120  
2121  
2122  
2123  
2124  
2125  
2126  
2127  
2128  
2129  
2130  
2131  
2132  
2133  
2134  
2135  
2136  
2137  
2138  
2139  
2140  
2141  
2142  
2143  
2144  
2145  
2146  
2147  
2148  
2149  
2150  
2151  
2152  
2153  
2154  
2155  
2156  
2157  
2158  
2159  
2160  
2161  
2162  
2163  
2164  
2165  
2166  
2167  
2168  
2169  
2170  
2171  
2172  
2173  
2174  
2175  
2176  
2177  
2178  
2179  
2180  
2181  
2182  
2183  
2184  
2185  
2186  
2187  
2188  
2189  
2190  
2191  
2192  
2193  
2194  
2195  
2196  
2197  
2198  
2199  
2200  
2201  
2202  
2203  
2204  
2205  
2206  
2207  
2208  
2209  
2210  
2211  
2212  
2213  
2214  
2215  
2216  
2217  
2218  
2219  
2220  
2221  
2222  
2223  
2224  
2225  
2226  
2227  
2228  
2229  
2230  
2231  
2232  
2233  
2234  
2235  
2236  
2237  
2238  
2239  
2240  
2241  
2242  
2243  
2244  
2245  
2246  
2247  
2248  
2249  
2250  
2251  
2252  
2253  
2254  
2255  
2256  
2257  
2258  
2259  
2260  
2261  
2262  
2263  
2264  
2265  
2266  
2267  
2268  
2269  
2270  
2271  
2272  
2273  
2274  
2275  
2276  
2277  
2278  
2279  
2280  
2281  
2282  
2283  
2284  
2285  
2286  
2287  
2288  
2289  
2290  
2291  
2292  
2293  
2294  
2295  
2296  
2297  
2298  
2299  
2300  
2301  
2302  
2303  
2304  
2305  
2306  
2307  
2308  
2309  
2310  
2311  
2312  
2313  
2314  
2315  
2316  
2317  
2318  
2319  
2320  
2321  
2322  
2323  
2324  
2325  
2326  
2327  
2328  
2329  
2330  
2331  
2332  
2333  
2334  
2335  
2336  
2337  
2338  
2339  
2340  
2341  
2342  
2343  
2344  
2345  
2346  
2347  
2348  
2349  
2350  
2351  
2352  
2353  
2354  
2355  
2356  
2357  
2358  
2359  
2360  
2361  
2362  
2363  
2364  
2365  
2366  
2367  
2368  
2369  
2370  
2371  
2372  
2373  
2374  
2375  
2376  
2377  
2378  
2379  
2380  
2381  
2382  
2383  
2384  
2385  
2386  
2387  
2388  
2389  
2390  
2391  
2392  
2393  
2394  
2395  
2396  
2397  
2398  
2399  
2400  
2401  
2402  
2403  
2404  
2405  
2406  
2407  
2408  
2409  
2410  
2411  
2412  
2413  
2414  
2415  
2416  
2417  
2418  
2419  
2420  
2421  
2422  
2423  
2424  
2425  
2426  
2427  
2428  
2429  
2430  
2431  
2432  
2433  
2434  
2435  
2436  
2437  
2438  
2439  
2440  
2441  
2442  
2443  
2444  
2445  
2446  
2447  
2448  
2449  
2450  
2451  
2452  
2453  
2454  
2455  
2456  
2457  
2458  
2459  
2460  
2461  
2462  
2463  
2464  
2465  
2466  
2467  
2468  
2469  
2470  
2471  
2472  
2473  
2474  
2475  
2476  
2477  
2478  
2479  
2480  
2481  
2482  
2483  
2484  
2485  
2486  
2487  
2488  
2489  
2490  
2491  
2492  
2493  
2494  
2495  
2496  
2497  
2498  
2499  
2500  
2501  
2502  
2503  
2504  
2505  
2506  
2507  
2508  
2509  
2510  
2511  
2512  
2513  
2514  
2515  
2516  
2517  
2518  
2519  
2520  
2521  
2522  
2523  
2524  
2525  
2526  
2527  
2528  
2529  
2530  
2531  
2532  
2533  
2534  
2535  
2536  
2537  
2538  
2539  
2540  
2541  
2542  
2543  
2544  
2545  
2546  
2547  
2548  
2549  
2550  
2551  
2552  
2553  
2554  
2555  
2556  
2557  
2558  
2559  
2560  
2561  
2562  
2563  
2564  
2565  
2566  
2567  
2568  
2569  
2570  
2571  
2572  
2573  
2574  
2575  
2576  
2577  
2578  
2579  
2580  
2581  
2582  
2583  
2584  
2585  
2586  
2587  
2588  
2589  
2590  
2591  
2592  
2593  
2594  
2595  
2596  
2597  
2598  
2599  
2600  
2601  
2602  
2603  
2604  
2605  
2606  
2607  
2608  
2609  
2610  
2611  
2612  
2613  
2614  
2615  
2616  
2617  
2618  
2619  
2620  
2621  
2622  
2623  
2624  
2625  
2626  
2627  
2628  
2629  
2630  
2631  
2632  
2633  
2634  
2635  
2636  
2637  
2638  
2639  
2640  
2641  
2642  
2643  
2644  
2645  
2646  
2647  
2648  
2649  
2650  
2651  
2652  
2653  
2654  
2655  
2656  
2657  
2658  
2659  
2660  
2661  
2662  
2663  
2664  
2665  
2666  
2667  
2668  
2669  
2670  
2671  
2672  
2673  
2674  
2675  
2676  
2677  
2678  
2679  
2680  
2681  
2682  
2683  
2684  
2685  
2686  
2687  
2688  
2689  
2690  
2691  
2692  
2693  
2694  
2695  
2696  
2697  
2698  
2699  
2700  
2701  
2702  
2703  
2704  
2705  
2706  
2707  
2708  
2709  
2710  
2711  
2712  
2713  
2714  
2715  
2716  
2717  
2718  
2719  
2720  
2721  
2722  
2723  
2724  
2725  
2726  
2727  
2728  
2729  
2730  
2731  
2732  
2733  
2734  
2735  
2736  
2737  
2738  
2739  
2740  
2741  
2742  
2743  
2744  
2745  
2746  
2747  
2748  
2749  
2750  
2751  
2752  
2753  
2754  
2755  
2756  
2757  
2758  
2759  
2760  
2761  
2762  
2763  
2764  
2765  
2766  
2767  
2768  
2769  
2770  
2771  
2772  
2773  
2774  
2775  
2776  
2777  
2778  
2779  
2780  
2781  
2782  
2783  
2784  
2785  
2786  
2787  
2788  
2789  
2790  
2791  
2792  
2793  
2794  
2795  
2796  
2797  
2798  
2799  
2800  
2801  
2802  
2803  
2804  
2805  
2806  
2807  
2808  
2809  
2810  
2811  
2812  
2813  
2814  
2815  
2816  
2817  
2818  
2819  
2820  
2821  
2822  
2823  
2824  
2825  
2826  
2827  
2828  
2829  
2830  
2831  
2832  
2833  
2834  
2835  
2836  
2837  
2838  
2839  
2840  
2841  
2842  
2843  
2844  
2845  
2846  
2847  
2848  
2849  
2850  
2851  
2852  
2853  
2854  
2855  
2856  
2857  
2858  
2859  
2860  
2861  
2862  
2863  
2864  
2865  
2866  
2867  
2868  
2869  
2870  
2871  
2872  
2873  
2874  
2875  
2876  
2877  
2878  
2879  
2880  
2881  
2882  
2883  
2884  
2885  
2886  
2887  
2888  
2889  
2890  
2891  
2892  
2893  
2894  
2895  
2896  
2897  
2898  
2899  
2900  
2901  
2902  
2903  
2904  
2905  
2906  
2907  
2908  
2909  
2910  
2911  
2912  
2913  
2914  
2915  
2916  
2917  
2918  
2919  
2920  
2921  
2922  
2923  
2924  
2925  
2926  
2927  
2928  
2929  
2930  
2931  
2932  
2933  
2934  
2935  
2936  
2937  
2938  
2939  
2940  
2941  
2942  
2943  
2944  
2945  
2946  
2947  
2948  
2949  
2950  
2951  
2952  
2953  
2954  
2955  
2956  
2957  
2958  
2959  
2960  
2961  
2962  
2963  
2964  
2965  
2966  
2967  
2968  
2969  
2970  
2971  
2972  
2973  
2974  
2975  
2976  
2977  
2978  
2979  
2980  
2981  
2982  
2983  
2984  
2985  
2986  
2987  
2988  
2989  
2990  
2991  
2992  
2993  
2994  
2995  
2996  
2997  
2998  
2999  
3000  
3001  
3002  
3003  
3004  
3005  
3006  
3007  
3008  
3009  
3010  
3011  
3012  
3013  
3014  
3015  
3016  
3017  
3018  
3019  
3020  
3021  
3022  
3023  
3024  
3025  
3026  
3027  
3028  
3029  
3030  
3031  
3032  
3033  
3034  
3035  
3036  
3037  
3038  
3039  
3040  
3041  
3042  
3043  
3044  
3045  
3046  
3047  
3048  
3049  
3050  
3051  
3052  
3053  
3054  
3055  
3056  
3057  
3058  
3059  
3060  
3061  
3062  
306

1590	Answer Choices:	1606	Q: What can an injury happen to?
1591	(a) freeway	1607	Answer Choices:
1592	(b) radio	1608	(a) emergency room
1593	(c) wires	1609	(b) doctor's office
1594	(d) computer network	1610	(c) living thing
1595	(e) electrical circuit	1611	(d) bones
1596	A: The answer must be a way to share files. The answer is	1612	(e) accident
1597	computer network (d).	1613	A: The answer must be something that can happen to a living
1598	A: The answer must be a connection which allows you to	1614	thing. The answer is living thing (c).
1599	share files. A computer network is a collection of	1615	A: The answer must be something that can happen to a living
1600	computers which are connected to each other. Therefore, the	1616	thing. If someone is injured, they are likely to want to
1601	answer is computer network (d).	1617	go to the emergency room. Therefore, the answer is living
1602	A: Files can be transmitted over a computer network, and	1618	thing (c).
1603	not a freeway. "wires" doesn't make any sense. Therefore,	1619	A: Living things feel pain, and therefore can be injured.
1604	the answer is computer network (d).	1620	Therefore, the answer is living thing (c).
1605			

## 1621 D Example Rationalizations on CQA

1622 We include a randomly sampled set of rationalizations which the model is able to produce before  
1623 fine-tuning. We observe that sometimes, the model constructs an argument roughly of the form “the  
1624 answer must have *a set of properties*. *correct answer* has *those properties*. therefore, the answer is  
1625 *correct answer*.” This structure of argument is fairly standard, but given that the model originally  
1626 answered those questions incorrectly, it resembles template-matching more than reasoning. The  
1627 technique of rephrasing the question in a simpler way before answering it directly is also often  
1628 effective. In general, the rationales from the pretrained model’s rationalization appears to be of  
1629 similar quality to the pretrained model’s rationale generation. We reference the pretrained model here  
1630 rather than the STaR-trained model, as the final iteration’s rationales are the ones which the model  
1631 struggled to learn from – that is, if the rationales were good, then we might expect that the model  
1632 would have already incorporated them into its rationale generation.

1633 Q: Why do I eat just enough when eating breakfast?

1634 Answer Choices:

- 1635 (a) gas
- 1636 (b) full stomach
- 1637 (c) feeling satisfied (CORRECT)
- 1638 (d) have energy
- 1639 (e) bloated

1640 A: The answer must be a reason for eating just enough when eating breakfast. Eating  
1641 just enough is a way to feel satisfied. Therefore, the answer is feeling satisfied (c).  
1642

1643

1644 Q: How is a dog likely to communicate with another dog?

1645 Answer Choices:

- 1646 (a) reproduce
- 1647 (b) bark (CORRECT)
- 1648 (c) do many things
- 1649 (d) clone
- 1650 (e) jump up

1651 A: The answer must be a way for a dog to communicate with another dog. Dogs bark to  
1652 communicate with each other. Therefore, the answer is bark (b).  
1653

1654 Q: If a person is trying to weasel out of punishment for a crime, where are they  
1655 likely to be?

1656 Answer Choices:

- 1657 (a) court room (CORRECT)
- 1658 (b) cherry tree
- 1659 (c) chicken coop
- 1660 (d) natural history museum
- 1661 (e) jail

1662 A: The answer must be a place where a person would be trying to get out of  
1663 punishment for a crime. The answer is court room (a).  
1664

1665 Q: What would easily hide a snake?

1666 Answer Choices:

- 1667 (a) living room of a house
- 1668 (b) kew gardens
- 1669 (c) terrarium
- 1670 (d) thick forest (CORRECT)
- 1671 (e) tropical forest

1672 A: The answer must be something that would hide a snake. The answer is thick forest  
1673 (d).  
1674

1675 Q: The person didn’t like the varying size of the cake slices handed out, she demand  
1676 more what?

1677 Answer Choices:

- 1678 (a) fairness (CORRECT)
- 1679 (b) candies
- 1680 (c) compliments
- 1681 (d) well fed
- 1682 (e) arguments

1683 A: The answer must be a reason for the person to demand more cake. The person is not  
1684 happy with the size of the cake slices. Therefore, the answer is fairness (a).  
1685  
1686 Q: In America, the courthouse of any given town has a likely location, where is it?  
1687 Answer Choices:  
1688 (a) electrical circuit  
1689 (b) getting married  
1690 (c) capital city  
1691 (d) center of town (CORRECT)  
1692 (e) michigan  
1693 A: The answer must be a location where a courthouse is located. The capital city is  
1694 the location of the capital of a state. Therefore, the answer is center of town (d).  
1695  
1696 Q: Where might I find a bowl nestled among other bowls in the kitchen?  
1697 Answer Choices:  
1698 (a) refrigerator  
1699 (b) kitchen top  
1700 (c) cupboard (CORRECT)  
1701 (d) dishwasher  
1702 (e) apartment  
1703 A: The answer must be a place where a bowl is found. The kitchen is a place where  
1704 bowls are found. Therefore, the answer is cupboard (c).  
1705  
1706 Q: Which state has blue ridge mountains that are called Blue Mountain?  
1707 Answer Choices:  
1708 (a) appalachians  
1709 (b) virginia  
1710 (c) pennsylvania (CORRECT)  
1711 (d) north carolina  
1712 (e) montana  
1713 A: The answer must be a state with mountains. The answer is Pennsylvania (c).  
1714  
1715 Q: Where could you find many radio studio?  
1716 Answer Choices:  
1717 (a) radio station  
1718 (b) country  
1719 (c) clear area  
1720 (d) office building  
1721 (e) large city (CORRECT)  
1722 A: The answer must be a place where many radio studios are located. Radio studios  
1723 are used to broadcast radio programs. Therefore, the answer is large city (e).  
1724  
1725 Q: Where would someone bring you a cup?  
1726 Answer Choices:  
1727 (a) apartment  
1728 (b) closet  
1729 (c) restaurant (CORRECT)  
1730 (d) table  
1731 (e) party  
1732 A: The answer must be a place where someone would bring you a cup. A restaurant is a  
1733 place where people bring cups. Therefore, the answer is restaurant (c).

## 1734 E STaR Without Rationalization (Rationale Generation Bootstrapping)

1735 For convenience, we include the rationale generation bootstrapping algorithm alone here. However, it  
1736 is simply Algorithm 1, with the blue components removed.

---

### Algorithm 2 Rationale Generation Bootstrapping (STaR without rationalization)

---

**Input**  $M$ : a pretrained LLM; dataset  $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^D$  (w/ few-shot prompts)

- 1:  $M_0 \leftarrow M$  # Copy the original model
- 2: **for**  $n$  **in**  $1 \dots N$  **do** # Outer loop
- 3:    $(\hat{r}_i, \hat{y}_i) \leftarrow M_{n-1}(x_i) \quad \forall i \in [1, D]$  # Perform rationale generation
- 4:    $\mathcal{D}_n \leftarrow \{(x_i, \hat{r}_i, y_i) \mid i \in [1, D] \wedge \hat{y}_i = y_i\}$  # Filter rationales using ground truth answers
- 5:    $M_n \leftarrow \text{train}(M, \mathcal{D}_n)$  # Finetune the original model on the correct solutions - inner loop
- 6: **end for**

---

## 1737 F Case Study

Q: John wants to eat at an Indian restaurant in on top of the space needle, where would he have to go?  
Answer Choices:  
(a) reservation (b) curry  
(c) british town (d) seattle  
(e) large city  
A: The answer must be a place with a space needle. The space needle is a landmark in Seattle. Therefore, the answer is Seattle (d).

Q: I put a towel in a cardboard box for the kitten, I wanted it to have what?  
Answer Choices:  
(a) farmhouse (b) zoos (c) cold place  
(d) floor (e) warm place  
A: The answer must be a place where the kitten would be comfortable. The kitten would be comfortable in a warm place. Therefore, the answer is warm place (e).

Q: Needlepoint requires many accessories, what holds the work taut?  
Answer Choices:  
(a) sewing (b) collection of folk art  
(c) livingroom (d) frame (e) bathroom  
A: The answer must be something that holds the work taut. The work is taut because it is held in place. Therefore, the answer is frame (d).

(a) Problems where the model was able to solve questions which it had previously answered incorrectly.

Q: They prided themselves on being a wealth of knowledge, and that's why many chose to attend their what?  
Answer Choices:  
(a) book store (b) university (c) meeting  
(d) class (e) encyclopedia  
~~A: The answer must be a place where people go to learn about things. The answer is university (b).~~  
A: The answer must be a place where people go to learn about things. Universities are places where people go to learn about things. Therefore, the answer is university (b).

Q: You want a real estate agent that is straight forward with you, the last thing you want is for them to what a house's quality?  
Answer Choices:  
(a) cover up (b) offer price  
(c) misrepresent (d) lie (e) falsify  
~~A: The answer must be a way to determine the quality of a house. The answer is misrepresent (c).~~  
A: The answer must be something that would be a bad thing for the real estate agent to do. The real estate agent would be misrepresenting the house's quality if they were to lie about it. Therefore, the answer is misrepresent (c).

(b) Problems where the model was already able to answer them with few-shot reasoning, but was able to answer better after more iterations.

Figure 7: Case study on STaR. We highlight some problems which the model is not initially able to solve, as well as rationales which are initially correct but improve. We note that for CQA, one is tasked with selecting the “best” answer, where multiple answers may be technically correct.

## G Further Limitations and Impacts

**Bias** It is important to note that STaR is designed to amplify the reasoning leading to correct solutions on a dataset. Another implication is that if biases are “useful” in solving a dataset then they will be amplified. This is made worse by rationalization, as biased answers that the model may not naturally arrive at are, in a sense, pulled out of the model. The exact interaction between the bias in a dataset and the pre-existing bias learned by the model is unclear, and something to consider before real-world deployment of all large language models in general, but STaR-trained models in particular.

We find some encouraging initial results on this however: for questions where gender is not relevant, the model appears to disregard it in its explanation. For example: “Q: Where is a good place for a woman to store her sunglasses? → A: The answer must be a place where sunglasses are stored. Sunglasses are stored in purses. Therefore, the answer is purse (e).” or “Q: The only baggage the woman checked was a drawstring bag, where was she heading with it? → A: The answer must be a place where a drawstring bag is checked. The answer is airport (e).” We believe this question warrants a much more comprehensive study.

**Faithfulness** One important challenge with models which seek to be interpretable or provide explanations for their reasoning is that of faithfulness. While our primary emphasis is not on the explainability benefits that STaR may bring, there is a fundamental challenge around evaluating explanations and rationales: namely, faithfulness [37, 38]. [38] describe faithful explanations as those which “accurately [represent] the reasoning process behind the model’s prediction.” While STaR encourages the use of reasoning in rationales which leads the model to correct answers, it is difficult, if not impossible, to ensure that the rationales reflect the model’s internal processing. For example, it is straightforward to imagine the model implicitly selecting a particular answer immediately and then generating a rationale to justify that selected answer. This would allow a model to generate unbiased rationales while selecting answers in a biased way.

The fact that our model outperforms one fine-tuned to directly predict the answers, and ablation studies from papers such as [6] make it clear that the generation of a rationale before producing an answer non-trivially improves the model’s answer quality. However, it is difficult to evaluate the degree to which any particular answer’s rationale is faithful. However, we note that there problems are not unique to STaR, but are symptomatic of the difficulty of understanding large language models and in particular the rationales generated by large language models.

**Scale** Finally, we note there is no guarantee that our results would generalize to larger models. However, [39] and [40] suggest that the benefits of rationales increases with scale on numerous problems, which is perhaps reason for optimism. On the other hand, a limitation of STaR on small models is that in order for the first iteration of STaR to succeed, few-shot performance must be above chance. This implies that the initial model must be big enough to have some reasoning capabilities. For instance we found that GPT-2 was not able to bootstrap from few-shot reasoning in even the arithmetic domain. A further limitation is that settings with a high level of chance performance (e.g. binary decisions) yield many poor rationales, confounding the STaR approach. As discussed, filtering bad reasoning paired with correct answers remains an open question.

## H Hyperparameters

GPT-J is a 28-layer decoder-only transformer, with an embedding size of 1024, 16 attention heads of dimension 256, and an FFN hidden layer of size 16384. It was pre-trained on the Pile [41], with a vocabulary size of 50.4K

In general, unless otherwise stated, we use a batch size of 8 sequences, each of length 1024. We also use packing, namely, packing the shorter examples to form longer sequences (up to length 1024) to improve TPU utilization. We do not use weight decay, and we train and sample on a single TPU-v3 node. We performed a hyperparameter search over learning rates from  $10^{-7}$  to  $10^{-4}$  using the Adam optimizer [42]. We found that  $10^{-6}$  was consistently the best-performing learning rate.

## 1786 I GSM8K Few-shot Prompt

1787 We include the following few-shot prompts for GSM8K, based on the examples in [9].

1788 Q: Natalia sold clips to 48 of her friends in April, and then she sold half as many  
1789 clips in May. How many clips did Natalia sell altogether in April and May?

1790 A: Natalia sold  $48/2 = \ll 48/2=24 \gg 24$  clips in May.

1791 Natalia sold  $48+24 = \ll 48+24=72 \gg 72$  clips altogether in April and May.

1792 ##### 72

1793

1794 Q: Betty is saving money for a new wallet which costs \$100. Betty has only half of  
1795 the money she needs. Her parents decided to give her \$15 for that purpose, and her  
1796 grandparents twice as much as her parents. How much more money does Betty need to  
1797 buy the wallet?

1798 A: In the beginning, Betty has only  $100 / 2 = \$\ll 100/2=50 \gg 50$ .

1799 Betty's grandparents gave her  $15 * 2 = \$\ll 15*2=30 \gg 30$ .

1800 This means, Betty needs  $100 - 50 - 30 - 15 = \$\ll 100-50-30-15=5 \gg 5$  more.

1801 ##### 5

1802

1803 Q: Julie is reading a 120-page book. Yesterday, she was able to read 12 pages and  
1804 today, she read twice as many pages as yesterday. If she wants to read half of the  
1805 remaining pages tomorrow, how many pages should she read?

1806 A: Maila read  $12 * 2 = \ll 12*2=24 \gg 24$  pages today.

1807 So she was able to read a total of  $12 + 24 = \ll 12+24=36 \gg 36$  pages since yesterday.

1808 There are  $120 - 36 = \ll 120-36=84 \gg 84$  pages left to be read.

1809 Since she wants to read half of the remaining pages tomorrow, then she should read  
1810  $84/2 = \ll 84/2=42 \gg 42$  pages.

1811 ##### 42

1812

1813 Q: Mark has a garden with flowers. He planted plants of three different colors in it  
1814 . Ten of them are yellow, and there are 80% more of those in purple. There are only  
1815 25% as many green flowers as there are yellow and purple flowers. How many flowers  
1816 does Mark have in his garden?

1817 A: There are  $80/100 * 10 = \ll 80/100*10=8 \gg 8$  more purple flowers than yellow flowers.

1818 So in Mark's garden, there are  $10 + 8 = \ll 10+8=18 \gg 18$  purple flowers.

1819 Purple and yellow flowers sum up to  $10 + 18 = \ll 10+18=28 \gg 28$  flowers.

1820 That means in Mark's garden there are  $25/100 * 28 = \ll 25/100*28=7 \gg 7$  green flowers.

1821 So in total Mark has  $28 + 7 = \ll 28+7=35 \gg 35$  plants in his garden.

1822 ##### 35

1823

1824 Q: Alexis is applying for a new job and bought a new set of business clothes to wear  
1825 to the interview. She went to a department store with a budget of \$200 and spent  
1826 \$30 on a button-up shirt, \$46 on suit pants, \$38 on a suit coat, \$11 on socks, and  
1827 \$18 on a belt. She also purchased a pair of shoes, but lost the receipt for them.  
1828 She has \$16 left from her budget. How much did Alexis pay for the shoes?

1829 A: Let S be the amount Alexis paid for the shoes.

1830 She spent  $S + 30 + 46 + 38 + 11 + 18 = S + \ll +30+46+38+11+18=143 \gg 143$ .

1831 She used all but \$16 of her budget, so  $S + 143 = 200 - 16 = 184$ .

1832 Thus, Alexis paid  $S = 184 - 143 = \$\ll 184-143=41 \gg 41$  for the shoes.

1833 ##### 41

1834

1835 Q: Tina makes \$18.00 an hour. If she works more than 8 hours per shift, she is  
1836 eligible for overtime, which is paid by your hourly wage + 1/2 your hourly wage. If  
1837 she works 10 hours every day for 5 days, how much money does she make?

1838 A: She works 8 hours a day for \$18 per hour so she makes  $8*18 = \$$

1839  $\ll 8*18=144.00 \gg 144.00$  per 8-hour shift

1840 She works 10 hours a day and anything over 8 hours is eligible for overtime, so she  
1841 gets  $10-8 = \ll 10-8=2 \gg 2$  hours of overtime

1842 Overtime is calculated as time and a half so and she makes \$18/hour so her overtime  
1843 pay is  $18*.5 = \$\ll 18*.5=9.00 \gg 9.00$

1844 Her overtime pay is  $18+9 = \$\ll 18+9=27.00 \gg 27.00$

1845 Her base pay is \$144.00 per 8-hour shift and she works 5 days and makes  $5 * \$144 = \$$   
1846  $\ll 144*5=720.00 \gg 720.00$

1847 Her overtime pay is \$27.00 per hour and she works 2 hours of overtime per day and  
1848 makes  $27*2 = \$\ll 27*2=54.00 \gg 54.00$  in overtime pay

1849 2 hours of overtime pay for 5 days means she makes  $54 \times 5 = \$270.00$   
1850 In 5 days her base pay is \$720.00 and she makes \$270.00 in overtime pay so she makes  
1851  $\$720 + \$270 = \$\langle\langle 720+270=990.00 \rangle\rangle 990.00$   
1852 ##### 990

## 1853 J STaR GSM8K Solutions

1854 We observe some interesting patterns with the GSM8K solutions proposed by the STaR-trained  
 1855 model. Typically, when the solution takes substantially fewer calculation steps than the ground truth,  
 1856 it corresponds to an instance where the model accidentally answered the question correctly despite  
 1857 mistakes in its reasoning. In some cases, however, the model produces simpler solutions than those in  
 1858 the ground truth. One example is shown in Figure 8.

Q: A van is delivering 180 bottles of drinks to a neighborhood, and each bottle contains either cider or beer or a mixture of two. Out of the 180 bottles, 40 contain only cider, 80 contain only beer, and the rest are a mixture of the two drinks. If the delivery man gives half the number of each bottle of drink to the first house, how many bottles does the first house get?	
<b>Human (Ground truth from GSM8K dataset):</b>  The total number of bottles containing cider or beer is $40+80 = \ll 40+80=120 \gg 120$ . If the number of bottles in the van was 180, then $180-120 = \ll 180-120=60 \gg 60$ contains a mixture of the two drinks. At the first house, the delivery man dropped $1/2 \cdot 40 = \ll 1/2 \cdot 40=20 \gg 20$ bottles. There were also $1/2 \cdot 80 = \ll 1/2 \cdot 80=40 \gg 40$ bottles of beer that the delivery man gave to the first house. Together, the first house received $40+20 = \ll 40+20=60 \gg 60$ bottles of cider and beer. The total number of bottles that had a mixture of the two drinks that were dropped at the first house is $1/2 \cdot 60 = \ll 1/2 \cdot 60=30 \gg 30$ . The first house got $60+30 = \ll 60+30=90 \gg 90$ bottles of all types of drinks. ##### 90	<b>STaR (on GPT-J):</b>  The delivery man gives half the number of each bottle of drink to the first house. So he gives $180/2 = \ll 180/2=90 \gg 90$ bottles of drink to the first house. ##### 90

Figure 8: An example problem in the training set where STaR derives a significantly simpler solution than the ground truth.